

Patricia M. French
Senior Attorney



300 Friberg Parkway
Westborough, Massachusetts 01581
(508) 836-7394
(508) 836-7039 (facsimile)
pfrench@nisource.com

July 20, 2005

BY OVERNIGHT DELIVERY AND E-FILE

Mary L. Cottrell, Secretary
Department of Telecommunications and Energy
One South Station
Boston, MA 02110

Re: Bay State Gas Company, D.T.E. 05-27

Dear Ms. Cottrell:

Enclosed for filing, on behalf of Bay State Gas Company ("Bay State"), please find Bay State's responses to the following Record Requests:

From the Attorney General:

RR-AG-25 RR-AG-37 RR-AG-47 RR-AG-48 RR-AG-54

From the Department:

RR-DTE-24 RR-DTE-38 RR-DTE-44 RR-DTE-46 RR-DTE-47 RR-DTE-59
RR-DTE-60 RR-DTE-61 RR-DTE-62 RR-DTE-63 RR-DTE-64 RR-DTE-65
RR-DTE-69 RR-DTE-70

From the USWA:

RR-USWA-1 RR-USWA-2 RR-USWA-4 RR-USWA-5 RR-USWA-6 RR-USWA-7
RR-USWA-8 RR-USWA-14

From the UWUA:

RR-UWUA-2

Please do not hesitate to telephone me with any questions whatsoever.

Very truly yours,

Patricia M. French

cc: Per Ground Rules Memorandum issued June 13, 2005:

Paul E. Osborne, Assistant Director – Rates and Rev. Requirements Div. (1 copy)

A. John Sullivan, Rates and Rev. Requirements Div. (4 copies)

Andreas Thanos, Assistant Director, Gas Division (1 copy)

Alexander Cochis, Assistant Attorney General (4 copies)

Service List (1 electronic copy)

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE ATTORNEY GENERAL
D.T.E. 05-27

Date: July 20, 2005

Responsible: John Skirtich, Consultant (Revenue Requirements)

RR-AG-25: Provide the dollar amount that was paid out by the company and included in the test-year cost of service for the judgments and settlements listed in AG-1-79, supplemental response.

Response: The amounts shown in the third column represent one-fifth of the amounts paid by the Company. In this filing, the Company proposed a 5-year average normalization adjustment for injuries and damages.

Incident Number	Amount Paid By Company	Amount Included in Test Year Cost of Service	Explanation
1	\$0	\$0	Bay State Gas Company was not a party in this proceeding
2	\$0	\$0	Bay State Gas Company was not a party in this proceeding
3	\$0	\$0	Bay State Gas Company was not a party in this proceeding
4	\$0	\$0	Bay State Gas Company was not a party in this proceeding
5	\$0	\$0	Bay State Gas Company was not a party in this proceeding
6	\$0	\$0	Settlement paid by co-defendant
7	\$17,500	\$3,500	Amount paid in 2004, in test year
8	\$0	\$0	Settlement paid by co-defendant
9	\$9,000	\$1,800	Amount paid in 2004, in test year
10	\$50,000	\$10,000	Amount paid in 2003, not in test year
11	\$1,000,000	\$200,000	Amount paid in 2003, not in test year
12	\$0	\$0	Settlement paid by co-defendant
13	\$0	\$0	Bay State Gas Company was not a party in this proceeding
14	\$0	\$0	Bay State Gas Company was not a party in this proceeding
15	\$13,255	\$0	Workers compensation self-insurance was

Incident Number	Amount Paid By Company	Amount Included in Test Year Cost of Service	Explanation
			eliminated in this case due to full insurance coverage.
16	\$0	\$0	Still in process
17	\$4,000	\$0	Amount paid in 2005, not in test year

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE ATTORNEY GENERAL
D.T.E. 05-27

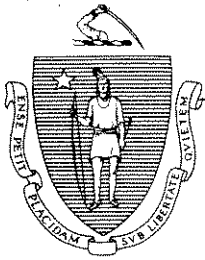
Date: July 20, 2005

Responsible: Stephen H. Bryant

RR-AG-37: Provide copies of the Department's letter dated October 20, 1999 and the Company's response letter dated April 14, 2000, that are referenced in the attachment to DTE-05-30.

Response: Attachment RR-AG-37 (a) is a copy of the requested letter dated October 20, 1999.

Attachment RR-AG-37 (b) is a copy of the requested letter dated April 24, 2000. The Company was unable to locate the attachments that are referenced in this letter.



THE COMMONWEALTH OF MASSACHUSETTS
OFFICE OF CONSUMER AFFAIRS AND BUSINESS REGULATION

Bay State Gas Company
D.T.E. 05-27
Attachment RR-AG-37 (a)
Page 1 of 3

DEPARTMENT OF
TELECOMMUNICATIONS & ENERGY

ONE SOUTH STATION
BOSTON, MA 02110
(617) 305-3500

ARGEO PAUL CELLUCCI
GOVERNOR

JANE SWIFT
LIEUTENANT GOVERNOR

DANIEL A. GRABAUSKAS
DIRECTOR OF CONSUMER AFFAIRS
AND BUSINESS REGULATION

JANET GAIL BESSER
CHAIR

JAMES CONNELLY, ESQ.
COMMISSIONER

W. ROBERT KEATING
COMMISSIONER

EUGENE J. SULLIVAN, JR.
COMMISSIONER

PAUL B. VASINGTON
COMMISSIONER

October 20, 1999

Mr. Richard P. Cencini
Vice President, Regulatory Affairs
Bay State Gas Company
300 Friberg Parkway
Westborough, MA 01581-5039

RE: Bay State Gas Company's Service Business

Dear Mr. Cencini:

This letter responds to Bay State Gas Company's ("Bay State" or "Company") proposal to the Department of Telecommunications and Energy ("Department") to retain its service business integrated within its corporate structure and utility distribution operations. Bay State identifies its service business to include furnace repairs, water heater repairs, water heater and conversion burner rentals, furnace inspections, and furnace installations (Cencini letter dated June 22, 1999). In the past, either Bay State or its affiliate, Energy USA, provided these services (id.). Bay State proposes to continue offering these services to its customers through its regulated utility business arguing that such provision allows Bay State to reduce distribution rates, avoid customer confusion, and better manage personnel workload (id.).

The Department is committed to bringing the benefits of competition to all utility customers. Electric Industry Restructuring, D.P.U. 96-100 (1996); Standards of Conduct, D.T.E. 96-44 (1996); Revised Standards of Conduct, D.T.E. 97-96 (1997); NOI - Natural Gas Unbundling, D.T.E. 98-32 (1998); Department of Telecommunications and Energy, 1998 Annual Report 2-3 (1999). As the Department has stated on numerous occasions, competition can be the best consumer protection and will lead to new and more efficient services and lower prices for unbundled services due to the advances and changes in the underlying economics and technologies of the utility industries. See, e.g., Department of Telecommunications and Energy, 1998 Annual Report 2-3 (1999); NOI - Natural Gas Unbundling, D.T.E. 98-32-B at 4-7 (1999). Bay State has acknowledged that "customers will benefit from competition." Status Report On The Massachusetts Gas Unbundling Collaborative, Att. E at 1 (March 18, 1998).

Mr. Richard P. Cencini
October 20, 1999

As the Department has stated in the past, for the full benefits of competition to accrue to consumers, the prerequisites of a true competitive market must be in place, including (1) many buyers and sellers with effective access to each other, (2) arm's length transactions between buyers and sellers, (3) broad and equal access to timely information, and (4) low thresholds for entry. Standards of Conduct, D.P.U. 96-44, at 2, Order Commencing Rulemaking (August 16, 1996). Most important, it is critical that no market participant, or group of participants, is in a position to exert unfair or abusive power in a new competitive industry structure. Id.

The Department recognizes that a natural gas local distribution company is in a position and has an incentive to exert undue preference and favorable treatment when providing services other than gas transportation and believes that this can be a serious threat to the development of a competitive marketplace. Corporate separation of these services provides an effective solution to the problem of anti-competitive transactions and requires less regulatory supervision than if the services remain integrated. See, Standards of Conduct, D.P.U. 96-44, Order Commencing Rulemaking, at 5 (August 16, 1996).

Bay State's proposal to remain, or to return to being, an integrated utility with respect to its service business is inconsistent with the Department's stated goals. As an integrated utility supplying both monopoly and competitive services, it also gives Bay State an opportunity and, perhaps, incentive and ability to discriminate in the provision of monopoly services in favor of its own competitive services. Separation of Bay State's service business from its regulated distribution services will reduce the possibility that regulated monopoly service might subsidize or unfairly discriminate in favor of the unregulated services. Separation will reduce the potential for inappropriate competitive advantage that Bay State might have by leveraging its monopoly services to its advantage in the otherwise competitive service business.

The Department has no reason to conclude that Bay State's service business is or should be exempt from the Department's regulatory policies on the relationship of a utility's monopoly and competitive services. See, e.g., Revised Standards of Conduct, D.T.E. 97-96, at 11-12 (1997); Standards of Conduct, D.T.E. 96-44, at 1, 6-7 (1996). We are not convinced that Bay State's claimed benefits to its distribution customers outweigh the potential harm from unintended subsidization of, or undue preference given to, the utility's competitive services.

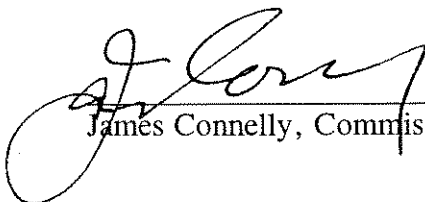
For the reasons stated in this letter, the Department does not approve Bay State's proposal to retain its service business integrated within its corporate structure and utility distribution operations. The Department directs Bay State either to establish its service

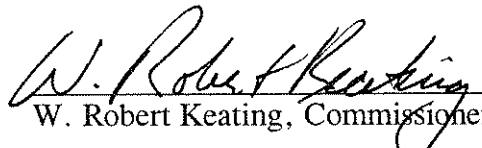
Mr. Richard P. Cencini
October 20, 1999

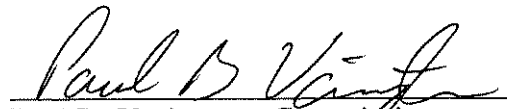
business as a separate unit within the Company, or to establish its service business outside its corporate structure as a separate legal entity. Under either option, Bay State and its service business are, and shall be, subject to the Department's Standards of Conduct, 220 C.M.R. 12.00 et seq.

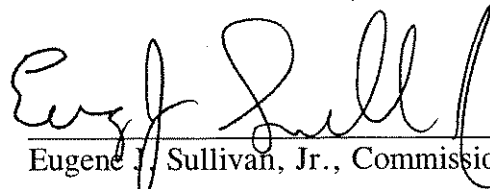
By Order Of The Department


Janet Gail Besser, Chair


James Connelly, Commissioner


W. Robert Keating, Commissioner


Paul B. Vasington, Commissioner


Eugene J. Sullivan, Jr., Commissioner

RUBIN AND RUDMAN LLP

COUNSELLORS AT LAW

50 ROWES WHARF • BOSTON, MASSACHUSETTS 02110-3319

TELEPHONE: (617) 330-7000 • FACSIMILE: (617) 439-9556 • EMAIL: FIRM@RUBINRUDMAN.COM

John A. DeTore
Direct Dial: (617) 330-7144
E-mail: jdetore@rubinrudman.com

April 14, 2000

Chairman James Connelly
Commissioner Paul Vasington
Commissioner Robert Keating
Commissioner Eugene Sullivan
Department of Telecommunications and Energy
One South Station, 2nd floor
Boston, Massachusetts 02110

Re: Bay State Gas Company's Service Business

Dear Commissioners:

On behalf of Bay State Gas Company ("Bay State" or the "Company"), I am writing in response to the Department of Telecommunications and Energy's ("Department" or "DTE") October 20, 1999 letter to Bay State concerning future treatment of Bay State's service business.

It is our understanding that the Department's October 20, 1999 letter voiced two major areas of concern regarding Bay State's existing service business model. First, the Department sought to ensure that costs and revenues associated with service business activities are properly accounted for to be certain that Bay State's provision of these services is not subsidized by its distribution ratepayers. Second, the Department expressed some concern regarding competitive implications of Bay State providing certain services, and, more specifically, whether Bay State actively provides access to its customers to non-affiliated contractors to perform these services.

In response to the Department's letter, Bay State has undertaken a comprehensive evaluation of its service business activities. Based on this evaluation, Bay State is prepared to enhance service business practices and procedures that are already in place to ensure that Bay State's participation strengthens competition and customer benefits. The remaining sections of this letter outline Bay State's proposal and describe how the proposal meets each of the Department's objectives.

I. SUMMARY OF PROPOSAL

Bay State recognizes that Department policy only requires that above the line activities such as the service business are profitable on an incremental basis. However, to allay all concerns about potential cross-subsidization, Bay State will track service business profitability

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 2

on the more stringent fully allocated basis. The procedures that Bay State has established to track fully allocated costs and revenues associated with the service business are described in greater detail below. In addition, Bay State is providing in this letter a detailed description of Bay State's current and enhanced contractor participation programs that address concerns about potential undue competitive advantage that Bay State may enjoy in terms of access to service business customers. ✓

II. OVERVIEW OF BAY STATE'S SERVICE BUSINESS

As the Department is aware, Bay State must be prepared as part of its obligation to provide safe and reliable service to its distribution customers, to maintain an infrastructure and trained personnel that are capable of responding to emergency customer calls, including odor and leak response services. In addition, Bay State currently offers the following services to customers on a fully integrated basis with utility operations: heating equipment repairs, water heater repairs, water heater and conversion burner rentals, heating equipment inspections, and heating equipment installations. Although gas distribution is a monopoly function, Bay State competes with oil and propane dealers plus regulated electric distribution utilities in the highly competitive "space heating" and "water heating" markets. The markets within which Bay State offers these services are highly competitive. Further, Bay State's overall role and market share in this service business area are vastly different from its role as provider of distribution service. For example, Bay State's share of heating system installation jobs along its distribution system is only 3% of the total market for this activity thus it cannot be presumed that Bay State exerts market power in these activities. ✓

Bay State views its ability to continue to offer these services as critical to several key objectives. First, continued operation of service business activities allows Bay State to maximize efficient use of existing resources, including union employees performing these services. As noted above, Bay State must keep in place the necessary personnel and infrastructure to provide safe and reliable service. Similarly, Bay State must maintain the appropriate levels of available resources to respond to emergency situations. Personnel performing service business activities undergo cross-training in all aspects of safety and code related functions. As a result, service technicians are capable of conducting visual safety checks and detecting gas odors or other potential safety problems during the normal course of business. By maintaining service business technicians within Bay State, these resources remain available to provide reliability and safety-related services, for the benefit of ratepayers. ✓

Second, Bay State uses these services as part of its overall efforts to attract new customers and to increase load from its existing customers. Increased load growth allows Bay State, in turn, to serve its existing customer base more efficiently and cost-effectively. Bay State's long-term growth strategy is designed in part to allow the Company to minimize the need for frequent rate increases in the future, thus benefiting existing ratepayers. Further, Bay State views its ability to offer these services as essential to the Company's ability to compete effectively with full service oil dealers, Bay State's primary competitors in the residential space. ✓

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 3

heating market. The ability to compete with unregulated oil dealers is significant not only with respect to Bay State's customer growth objectives, but also to meeting a third key objective, satisfying Bay State's existing customer base. Both market research efforts and experience with customers have demonstrated to Bay State that many customers want and expect to be able to obtain rental or repair services from Bay State. Accordingly, Bay State considers its continued ability to provide these types of services to existing and prospective customers as critical to maintaining a high level of customer satisfaction.

Given the Company's growth strategy, Bay State looks to independent contractors as trade allies who can assist the Company in its efforts to attract new load and to increase load from existing customers. Consequently, as discussed in greater detail in Section IV, *infra*, Bay State makes significant efforts to solicit participation by qualified contractors in its contractor referral program and provides incentives to participants in the program to help meet company growth objectives. It also bears noting that Bay State's experience has been that customer demand for certain of the activities it provides to customers, such as conversion burner rentals, simply would not be met if left exclusively to the independent contractors.

Bay State recognizes that specific business objectives as well as the strategies implemented to meet those goals will invariably differ from utility to utility and that some individual flexibility in approach regarding service business activities is warranted. However, Bay State believes that the particular approach outlined here is appropriate given the specific circumstances of Bay State's stated business objectives, existing operating structure and the demonstrated benefits to customers.

III. PREVENTION OF CROSS-SUBSIDIZATION

A. Introduction

The types of services offered under the umbrella of the "service business" are core to the provision of natural gas distribution service; however, neither the services offered nor the prices and terms under which Bay State competes in this market are regulated by the Department.

Historically, under traditional cost-of-service ratemaking, DTE precedent called for the careful examination of non-utility business activities, such as appliance rentals, whether treated above-the-line or below-the-line for ratemaking purposes. Where the non-utility service is above-the-line for ratemaking purposes (i.e., costs and revenues are included in the utility's overall revenue requirement), the DTE has utilized an incremental cost approach, to ensure that: (1) ratepayers do not subsidize these separate business activities, and (2) ratepayers share in any benefits attributable to such activities. Because Bay State included these types of services "above-the-line"¹ for ratemaking purposes, the Department has previously required a comparison

¹ Service business activities have been treated above-the-line with one exception: sales and installations are below-the-line for ratemaking purposes. Therefore, costs and revenues historically have been removed from Bay State's cost of service.

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 4

of incremental revenues and costs attributable to the specific activity, to ensure that ratepayers do not subsidize such activities. See, e.g., Bay State Gas Company, D.P.U. 89-81 (1989); Bay State Gas Company, D.P.U. 92-111 (1992). In D.P.U. 89-81, the Department found:

the proper accounting treatment for an above-the-line activity is an incremental approach. With an incremental approach, the ratepayers experience the benefit of any incremental profitability of a program. If, using an incremental approach, an above-the-line activity did not produce profits, the Department could take steps necessary to ensure that ratepayers are protected.

D.P.U. 89-81, at 73, citing Commonwealth Gas Company, D.P.U. 87-122 at 21 (1987).

In general, DTE precedent favors allocation of costs between utility and non-utility uses based on the same method of allocation. Further, the DTE will look to ensure that corresponding adjustments (e.g., depreciation reserve associated with plant allocated to the non-utility business) are also made. Finally, where possible, the DTE expects that costs attributable to activities like these will be directly assigned.

Bay State is ready to utilize the following steps to track service business costs and revenues on a fully allocated basis. By tracking service business costs on a fully allocated basis rather than following the Department's incremental cost precedent, Bay State will be meeting a more rigorous standard, thereby ensuring beyond any reasonable doubt that the regulated monopoly services are not subsidizing Bay State's competitive service activities.

B. Cost Allocation Methodology

Bay State proposes to track the costs associated with the service business activities in the following manner:

- Direct Costs will include direct labor, parts & materials, rental water heater lease expense and rental water heater and conversion burners depreciation expense. These costs are charged directly to the service activities as incurred.
- Direct Fringes will include company benefits, payroll taxes and liability insurance related to direct labor. The costs are allocated by multiplying direct labor by the percentage of total company fringes to total company payroll (the "fringe benefit rate").
- Overhead will include the following costs which are directly charged or allocated between the total service activities and other utility activities on the bases set forth in the chart below. The costs for total service activities are subsequently allocated among the individual service activities on the basis of direct labor.

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 5

Overhead Calculations	
Cost Item	Direct Charge/Allocation Basis
Indirect Payroll	
Supervision	Allocation- estimated time spent
Vacation, holiday, sickness, training, etc	Allocation-Direct labor
Dispatch	Allocation-Estimated time spent and no. of customers by state
Workforce Planning	Allocation-Estimated time spent
Sales	Allocation-Estimated time spent
Customer Call Center	Allocation-Type of customer call received and no. of customers by state
Administrative	Allocation-Estimated time spent
Indirect Fringes	Allocation-Indirect Payroll multiplied by fringe benefit rate
Tools & Equipment	Direct Charge-tools used
Uniform Rental	Direct Charge-uniforms used
Advertising	Direct Charge-specific programs
Bad Debts	Direct Charge-accounts uncollectible
Office Expense	Direct Charge-items purchased
Fleet Expense	Allocation-No. of vehicles and payroll dollars
Stores Expense	Allocation-Cost of materials withdrawn from M&S inventory
Facility Space	Allocation-Proportion of facility space used

IV. COMPETITIVE ISSUES

As previously noted, the Department's October 20, 1999 letter raised a concern that Bay State's participation in the service business could threaten development of a competitive marketplace, to the extent that favorable treatment could be utilized to advantage Bay State's service business. Letter at 2. Accordingly, the Department directed Bay State to take steps to ensure that the Department's stated concerns were met. *Id.* Bay State believes the proposed cost allocation tracking system described above will allow the Department to ensure that Bay State's service business activities are not subsidized by ratepayers and compete fairly with independent contractors in this regard. Moreover, as described below, Bay State already has implemented a robust contractor referral program to ensure that customers are aware of their options. In addition to providing leads to Bay State participating contractors, Bay State provides these contractors with rebates for eligible installation work and offers additional incentives and/or customers rebates through their participation in the program.

A. Overview of BSG Contractor Referral Efforts

A thorough review of Bay State's efforts to date to solicit independent contractors for its referral program demonstrates Bay State's commitment to benefiting customers through competition by treating contractors as allies or partners. Specifically, Bay State has undertaken

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 6

significant efforts to conduct outreach to independent contractors and to provide referrals to Bay State customers in need of particular services, and will continue to do so. The process outlined below was used to develop the 1999 contractor referral list; the process is similar from year to year and was used to develop the list for this year also.

1. Identification and Outreach Efforts directed at Independent Contractors

In August 1998, Bay State undertook a comprehensive effort to identify all potential contractors for its contractor referral list. For each of its operating divisions, Bay State reviewed and updated lists of plumbing and heating contractors operating in the service area. Sources for this information include prior Bay State outreach efforts, requests made by individual contractors to be added to Bay State's referral program, and state licensing board directory information. The information obtained resulted in a combined list of 585 Massachusetts HVAC contractors. In September 1998, Bay State sent an invitational mailing to all 585 identified contractors ("September 1998 Mailing"). As described in that document, Bay State offered the opportunity to all recipients to have their company names included in Bay State's Participating Contractor referral lists, which are provided to customers seeking referrals. A sample copy is included as Attachment A. The September 1998 Mailing invited all prospective 585 participants to attend dinner meetings, which were held in each of Bay State's operating divisions between November 2-5, 1998. As a result of elimination of duplicate names or businesses for whom mail was returned as undeliverable, the total list of prospective participants was subsequently reduced to 529. At each of those dinner meetings, Bay State provided to attendees an information packet explaining the calendar year 1999 program. Each packet included a program outline brochure, a copy of the presentation delivered by the Regional Sales leader, a blank participating contractor agreement,² and a service territory check sheet for the appropriate division. A copy of the information packet is included as Attachment B. The information provided at the dinner meetings also described all applicable bonus/incentive programs for which contractors could qualify. Bay State established a December 18, 1998 deadline for responses to the solicitation. However, this deadline was extended for various contractors who expressed interest, but had failed to complete the necessary documents. In addition, Bay State personnel conducted substantial follow up telephone calls to remind prospective participants to provide necessary information that had not been received. The end result of this comprehensive effort to identify all prospective independent contractors and conduct outreach efforts resulted in a total of 111 contractors completing the necessary information to qualify for the referral program. In February 1999, lists of referral contractors were printed and distributed to each of Bay State's operating divisions. Copies of the 1999 contractor referral lists are included as Attachment C.

In September 1999, Bay State again contacted participating contractors to update types of activities performed. After receipt of these responses, and review of the referral list by Bay State's sales, service, business improvement and call center areas to remove any contractors who

² The participant agreement form requests general information regarding types of services provided and geographic area covered, as well as information necessary to allow Bay State to ensure that the contractor is appropriately licensed and insured.

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 7

had failed to provide service when called upon or whose work proved to be of poor quality, the contractor referral list was updated. Updated lists for each service territory area were then prepared and delivered to each of Bay State's operating divisions.

2. Use of the Contractor Referral List

Bay State's Contractor Referral list is distributed in three primary ways: (1) Bay State's sales department provides the list to prospective conversion customers, unless the customer has already retained a contractor or is seeking installation of a conversion burner³. Similarly, Bay State's call center representatives are instructed to offer to customers the list of contractors who will install heating equipment; (2) The list is provided to customers when a Bay State employee "red tags" (i.e., finds a safety condition requiring immediate repair or replacement) an appliance at a particular location; (3) in addition, a separate referral list is used for service calls during peak periods for customer convenience and choice. When customers call for service during off-peak periods they will be informed that independent contractors also provide the same service, and that if the customer wants more information we can provide the names of these independent contractors to the customers over the phone. In addition to these general methods of distribution, any time a customer seeks information on contractor referrals from Bay State, the Company provides a copy of the referral list. The company will continually focus on communicating to customers that, along with having Bay State as a choice for their service provider, there are independent contractors available to provide this service. For example, Bay State is investigating the feasibility of posting the list on its web sites. Another possibility is to include information on contractors as an option on future automated services. Although Bay State utilizes the contractor referral process to the degree possible, in certain circumstances, e.g., emergency gas leak calls, the particular service must be performed by Bay State personnel.

V. CONCLUSION

As previously noted, Bay State believes that the measures outlined in this letter are responsive to the concerns voiced by the Department in its October 20, 1999 letter. In sum, Bay State submits that it has demonstrated that its distribution and service customers will share in the benefits derived from providing these services, including increased load growth and efficient use of existing resources. Moreover, market research shows that Bay State's customers want and expect the continued provision of these services through Bay State. At the same time, Bay State has developed measures that will fully insulate its ratepayers from all risks associated with the service business. Accordingly, Bay State believes its proposal is in the best interests of ratepayers.

Finally, we note that Bay State has attempted herein to address the Commission and DTE staff's concerns in full. Should any additional questions remain unanswered, or should the Commission or DTE staff require additional information concerning any of these activities, Bay

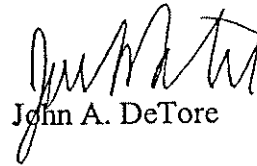
³ As a general rule, certain types of services, such as conversion burner and water heater rentals are not commonly available from independent contractors.

RUBIN AND RUDMAN LLP

Letter to DTE Commissioners
April 14, 2000
Page 8

State would be happy to respond to such requests. We look forward to hearing from you regarding this proposal.

Sincerely,



John A. DeTore

JAD/df

cc: Paul Afonso, Esq.
Jeffrey Yundt
Richard Cencini
George Yankos
Rebecca Hanson

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE ATTORNEY GENERAL
D.T.E. 05-27

Date: July 20, 2005

Responsible: John Skirtich, Consultant (Revenue Requirements)

RR-AG-47: Has the Company filed any lawsuits or joined in any lawsuits against insurance brokers regarding their brokerage services. If any lawsuits have been filed, has the company received any benefits from the lawsuits

Response: NiSource has not filed or joined in any lawsuits against any of its brokers.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE ATTORNEY GENERAL
D.T.E. 05-27

Date: July 20, 2005

Responsible: John E. Skirtich, Consultant (Revenue Requirements)

RR-AG-48: Identify if TriNet Corporate Realty Trust, Inc. was an affiliate of NiSource at any time from June 1, 1997 to the present.

Response: To the best of our knowledge, the only relationship with TriNet Corporate Realty Trust, Inc. and NiSource was by means of the sale lease back arrangement made prior to the merger of NIPSCO with BSG. Our records depicting a chronological time line since 1853 of all companies and organizations that eventually make-up the NiSource profile as we know it today do not indicate that TriNet Corporate Realty Trust, Inc. was a subsidiary, predecessor or had any other affiliate relationship with NiSource. We believe the "Whiting" name in Whiting Clean Energy and Whiting Electric Company (organized in 1896) are both more indicative of a geographic location, Whiting, Indiana, rather than Mr. Mark Whiting as shown on the sale lease-back contract for the Westborough facility.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE ATTORNEY GENERAL
D.T.E. 05-27

Date: July 20, 2005

Responsible: Stephen H. Bryant, President

RR-AG-54: Provide a copy of the letter sent by Bay State to the DTE Consumer Division, indicating the closing of Bay State's walk-in centers.

Response: A copy of the above referenced letter from Bay State to the DTE Consumer Division was provided in USWA 2-2, as Attachment USWA 2-2 (b).

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: John Skirtich, Consultant (Revenue Requirements)

RR-DTE-24: Is there a manual to describe the process related to DTE 9-21? If so, provide.

Response: Below is a list of the definitions of the acronyms used in Attachment DTE-9-21.

SONP – Shut Off Non Payment
CSR – Customer Service Representative
NSF – No Sufficient Funds

Please refer to the Company's response to AG-22-13 for more details concerning the process.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: John Skirtich, Consultant (Revenue Requirements)

RR-DTE-38: Reconcile calculated bad debt in Mr. Skirtich's schedules to the bad debt number on p. 9 of 2004 Annual Return, line 40. Provide similar information for 2003.

Response: Attachment RR-DTE-38 details by account the activity in the Reserve for Uncollectible Accounts (260) as shown in the Annual Return, Line 40 from 2004 and 2003. Notes are provided referencing amounts included in the Company's cost of service filing. During the review, it was discovered, that the 2004 charge-offs included a reclass of \$38,131 (see note a of the Attachment RR-DTE-38 for 2004). The Company will make the appropriate correction to the filing.

Bay State Gas Company
Reserve for Uncollectible Accounts (260)

Account No.	Description	2003 \$	Total Write-offs \$	Bad Debt Accrual \$	Reclass Tax Refund Adjustment \$	Reclass Unbilled \$	Bad Debt Gas Cost Acct 182-19 \$	Below the Line Acct 415-16 \$	2004 \$
260-01	Reserve-Gas Uncoll Acct	2,195,000	(9,038,394) (a)	9,549,524 (b)	(38,131) (a)	(223,000)	0	0	2,445,000
	Adjust expense to be non-gas cost portion	0	0	(6,595,000)	0	0	6,595,000	0	0
	Charge expense based on recoveries - CGA	0	0	5,290,135	0	0	(5,290,135)	0	0
260-02	Reserve-Merch Uncoll Acct	81,000	(17,905)	0	0	0	0	(44,095)	19,000
260-05	Reserve-Sundry Uncoll Acct	350,000	(117,092) (c)	61,092 (d)	0	0	0	0	294,000
260-09	Reserve-Water Heater Uncoll Acct	1,018,000	(451,029) (c)	326,029 (d)	0	0	0	0	893,000
260-11	Reserve-Guardian Care Uncoll Acct	162,000	(41,645) (c)	25,645 (d)	0	0	0	0	146,000
260-20	Reserve Unbilled	680,000	0	0	0	223,000	0	0	903,000
260-21	Reserve Special	0	0	12,300	0	0	0	0	12,300
260-22	Reserve Special F	1,925,000	0	232,869	38,131	0	604,000	0	2,800,000
	Total	6,411,000	(9,666,065)	8,902,595	0	0	1,908,865	(44,095)	7,512,300

(a) See Exh. BSG/JES-1, Workpaper JES-6, Page 21, Line 44.

(b) See Data Request DTE-9-1.

(c) See Exh. BSG/JES-1, Workpaper JES-6, Page 22, Line 42.

(d) See Exh. BSG/JES-1, Schedule JES-6, Page 10, Line 10

Account No.	Description	2002 \$	Total Write-offs \$	Bad Debt Accrual \$	Reclass \$	Reclass Unbilled \$	Bad Debt Gas Cost Acct 182-19 \$	Below the Line Acct 415-16 \$	2003 \$
260-01	Reserve-Gas Uncoll Acct	3,183,000	(8,727,904) (a)	8,420,000	0	(680,000)	0	0	2,195,096
	Adjust Acct 5182-19 from 56% to 65%			(238,850)			238,850		0
	Special large customer charge-offs handled outside normal channels		(1,208,383) (a)	407,225	424		800,638		(96)
	Adjust expense to be non-gas cost portion			(5,582,000)			5,582,000		0
	Charge expense based on recoveries - CGA			5,303,552			(5,303,552)		0
260-02	Reserve-Merch Uncoll Acct	2,946	(51,295)	0	0	0	0	129,349	81,000
260-03	Reserve-Propane Uncoll Acct	603	0		(603)	0	0	0	(0)
260-05	Reserve-Sundry Uncoll Acct	27,254	(378,674) (b)	701,844	(424)	0	0	0	350,000
260-09	Reserve-Water Heater Uncoll Acct	41,325	(329,972) (b)	1,306,044	603	0	0	0	1,018,000
260-11	Reserve-Guardian Care Uncoll Acct	17,462	(46,106) (b)	190,644	0	0	0	0	162,000
260-20	Reserve Unbilled	0	0	0	0	680,000	0	0	680,000
260-21	Reserve Special	136,000	0	(136,000)	0	0	0	0	0
260-22	Reserve Special F	0	0	674,000	0	0	1,251,000	0	1,925,000
	Total	3,408,590	(10,742,334)	11,046,459	0	0	2,568,936	129,349	6,411,000

(a) See Exh. BSG/JES-1, Workpaper JES-6, Page 21, Line 31.

(b) See Exh. BSG/JES-1, Workpaper JES-6, Page 22, Line 28.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: John Skirtich, Consultant (Revenue Requirements)

RR-DTE-44: D.T.E. 16-34, p. 1 of 4, item #4, breakdown of Acct. 257-03. Res. Amort. Org.

Response: Based on the Commission Order issued October 30, 1992 in Case No. D.P.U. 92-111, the Department noted that it had previously approved inclusion in rate base of the organization costs related to the mergers of Lawrence Gas Company, Brockton-Taunton Gas Company and Bay State Gas Company, the predecessor companies that now comprise Bay State's retail Massachusetts operations. Please see Page 68 of the Order.

Following this Order and the adjustments made in the 1992 case, the Company eliminated the same \$3,743,730 (as shown in D.P.U. 92-111, Workpaper BSG-3-5 and provided in D.T.E-16-34) from rate base and made a corresponding adjustment of \$2,936,755 to reserve for amortization related to the \$3,743,730.

The remaining \$152,920 in Account 257-03 reflects reserve associated with organization costs of \$195,536 included in Account 303 and included in rate base interpreted as being approved by the Commission in previous proceedings.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: Stephen H. Bryant

RR-DTE-46: Confirm the purpose of Form UCC-3 as referenced in the Company's response to Information Request DTE-01-20 at Attachment DTE-01-20 (a), page 13.

Response: The UCC-3 as referenced at page 13 of Attachment DTE-01-20 (a) was intended to more accurately describe the meter reading equipment that would be covered under the lease. The original UCC-1 described the equipment as 157,585 Itron residential monitoring devices. The UCC-3 more accurately describes the assets as 147,015 Itron residential monitoring devices and 10,570 commercial devices.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: Stephen H. Bryant

RR-DTE-47: Confirm the purpose of Pages 15 through 18 in the Company's response to Information Request DTE-01-20, Attachment A.

Response: The purpose of the above-referenced pages (various certificates of insurance) is to demonstrate to the bank/lessor that the Company has satisfied its contractual commitment to maintain certain types and levels of insurance during the lease term.

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: Lawrence R. Kaufmann, Consultant (PBR)

RR-DTE-59: Provide an update to DTE-4-10 that includes an earthquake dummy variable.

Response: See Attachment RR-DTE-59 for a copy of the updated model. The earthquake dummy is statistically significant in this specification. In the July 8, 2005 supplemented response to DTE-4-10, Bay State's actual costs are now 1.4% above their predicted value, and the difference is not statistically significant.

> run C:\Work\BayState\Specification\totalcost;

Bay State Gas Company

***** DTE 05-27
GAUSS Data Export Facility Attachment RR-DTE-59
***** Page 1 of 5

Begin export...

Export completed

Number of cases in GAUSS data set: 453.000

Number of cases written to foreign file : 453.000

Number of variables written to foreign file : 31.000

Date: 7/18/05 **** SUR ESTIMATION RESULTS **** Time: 14:18:44

OUTPUT FILE:C:\work\Baystate\results\dr_10a

DATA FILE:C:\work\Baystate\bench03dr10.xls

DEFINITIONS OF OUTPUT VARIABLES:

Y1 is number of customers.

Y2 is Total deliveries.

DEFINITIONS OF BUSINESS CONDITION VARIABLES:

Z1 is % of non-iron and steel in Dx miles

Z2 is Number of Electric Customers

Z3 is northeast dummy variable

Z4 is Miles of Distribution Main

Z5 is Pbr dummy variable for Bay State Gas

Z6 is Earthquake Dummy Variable

Model includes time trend.

Time period used: 1994 through 2003

k = nadd10

GAUSS Data Import Facility

Begin import...

Import completed

Number of rows in input file: 473

Number of cases written to GAUSS data set: 473

Number of variables written to GAUSS data set: 60

1
453

=====
SEEMINGLY UNRELATED REGRESSION WITH HETEROSKEDASTICITY 7/18/2005 2:18 pm
=====

Data Set: C:\work\Baystate\Temp_3.dat

Page 1 of 5

DIVISOR USING N IN EFFECT
RESTRICTIONS IN EFFECT

ITER. # =	0	LOG OF DETERMINANT OF SIGMA =	5.32079109
ITER. # =	1	LOG OF DETERMINANT OF SIGMA =	5.24991798
ITER. # =	2	LOG OF DETERMINANT OF SIGMA =	5.24689936
ITER. # =	3	LOG OF DETERMINANT OF SIGMA =	5.24659728
ITER. # =	4	LOG OF DETERMINANT OF SIGMA =	5.24656359
ITER. # =	5	LOG OF DETERMINANT OF SIGMA =	5.24655980
ITER. # =	6	LOG OF DETERMINANT OF SIGMA =	5.24655938
ITER. # =	7	LOG OF DETERMINANT OF SIGMA =	5.24655933
ITER. # =	8	LOG OF DETERMINANT OF SIGMA =	5.24655932
ITER. # =	9	LOG OF DETERMINANT OF SIGMA =	5.24655932
ITER. # =	10	LOG OF DETERMINANT OF SIGMA =	5.24655932

Equation: 1
Dependent variable: C

Total cases:	453	Valid cases:	453
Total SS:	388.703	Degrees of freedom:	----
R-squared:	0.966	Rbar-squared:	0.965
Residual SS:	13.056	Std error of est:	2.268
Durbin-Watson:	0.268		

Variable	Estimated Coefficient	Standard Error	t-ratio	Prob > t
CONST	8.08808246	0.01716690	471.144	0.0000
WL	0.20501219	0.00284975	71.940	0.0000
WK	0.63863075	0.00299650	213.126	0.0000
Y1	0.53553367	0.03910294	13.695	0.0000
Y2	0.24327566	0.03269202	7.441	0.0000
WLWL	0.03923189	0.03703039	1.059	0.2900
WLWK	-0.13223321	0.02507560	-5.273	0.0000
WKWK	0.21157355	0.02214708	9.553	0.0000
Y1Y1	-0.43356115	0.06221023	-6.969	0.0000
Y2Y2	-0.43520590	0.08609435	-5.055	0.0000
WLY1	-0.01908993	0.00726164	-2.629	0.0089
WLY2	-0.01324480	0.00722657	-1.833	0.0675
WKY1	0.01267853	0.00515329	2.460	0.0143
WKY2	0.01364242	0.00601348	2.269	0.0238
Y1Y2	0.41446033	0.07088294	5.847	0.0000
Z1	0.02498041	0.04620705	0.541	0.5890
Z2	-0.00601630	0.00100602	-5.980	0.0000
Z3	0.09803833	0.00738140	13.282	0.0000
Z4	0.09074231	0.03389723	2.677	0.0077
Z5	-0.00065876	0.00082948	-0.794	0.4275
Z6	0.02311896	0.00308579	7.492	0.0000
TREND	-0.01998242	0.00201065	-9.938	0.0000
K	-0.05089007	0.01100199	-4.626	0.0000

Equation: 2
Dependent variable: SL

Total cases:	453	Valid cases:	453
Total SS:	2.719	Degrees of freedom:	----
R-squared:	0.090	Rbar-squared:	0.102
Residual SS:	2.473	Std error of est:	2.524
Durbin-Watson:	0.368		

Variable	Estimated Coefficient	Standard Error	t-ratio	Prob > t
CONST	0.20501219	0.00284975	71.940	0.0000
WL	0.03923189	0.03703039	1.059	0.2900
WK	-0.13223321	0.02507560	-5.273	0.0000
Y1	-0.01908993	0.00726164	-2.629	0.0089
Y2	-0.01324480	0.00722657	-1.833	0.0675

Equation: 3
Dependent variable: SK

Total cases:	453	Valid cases:	453
Total SS:	3.369	Degrees of freedom:	----
R-squared:	0.183	Rbar-squared:	0.193
Residual SS:	2.754	Std error of est:	3.385
Durbin-Watson:	0.280		

Variable	Estimated Coefficient	Standard Error	t-ratio	Prob > t
CONST	0.63863075	0.00299650	213.126	0.0000
WL	-0.13223321	0.02507560	-5.273	0.0000
WK	0.21157355	0.02214708	9.553	0.0000
Y1	0.01267853	0.00515329	2.460	0.0143
Y2	0.01364242	0.00601348	2.269	0.0238

Equation: 4
Dependent variable: SM

Valid cases:	453
Degrees of freedom:	----

Variable	Estimated Coefficient	Standard Error	t-ratio	Prob > t
CONST	0.15635706	0.00269049	58.115	0.0000
WL	0.09300132	0.02339199	3.976	0.0001
WK	-0.07934034	0.01743502	-4.551	0.0000
Y1	0.00641140	0.00650746	0.985	0.3257
Y2	-0.00039762	0.00684489	-0.058	0.9537

MEASURES OF GOODNESS-OF-FIT

AN UNCENTERED SYSTEM R-SQUARE	0.968
A CENTERED SYSTEM R-SQUARE	0.971

The results from the test of the null hypothesis that all slope coefficients in all equations are simultaneously equal to zero.

Test statistic	Prob > t
1603.514	0.0000

 VALIDATION OF REGULARITY CONDITIONS

Monotonicity of the Estimated Cost Function

The number of observations for which each of the following predicted cost share is nonpositive is listed below

Labor	Capital	Materials
0	0	0
(0.00 %)	(0.00 %)	(0.00 %)

Concavity of the Estimated Cost Function

The number of the observations for which the condition that the matrix of second order partial derivatives of the cost function with respect to input wages is negative semi-definite holds:

423 (93.38 %)

Quasi-Concavity of the Estimated Cost Function

The number of observations for which the condition that the cost function is strictly quasi-concave in input prices holds:

423 (93.38 %)

Second Order Condition for Cost Minimization

The number of the observations for which the condition that the bordered Hessian is negative definite holds:

423 (93.38 %)

OUT-OF-SAMPLE PREDICTION OF TOTAL COST LEVEL PERFORMANCE LAST 5 YEARS

Actual	Predicted	Difference	t_ratio	p_value	Utility
7.375	7.716	-0.341	-11.917	0.000	30.000
7.810	8.141	-0.330	-11.536	0.000	38.000
8.231	8.499	-0.268	-9.344	0.000	15.000
6.765	7.028	-0.263	-9.185	0.000	53.000
7.566	7.771	-0.206	-7.181	0.000	44.000
8.538	8.694	-0.157	-5.473	0.000	12.000
8.844	8.999	-0.155	-5.404	0.000	23.000
9.739	9.880	-0.141	-4.918	0.000	40.000
7.893	8.033	-0.140	-4.873	0.000	46.000
5.993	6.118	-0.125	-4.359	0.000	7.000
7.255	7.372	-0.117	-4.089	0.000	37.000
6.607	6.720	-0.113	-3.952	0.000	26.000
7.639	7.750	-0.111	-3.870	0.000	49.000
8.063	8.163	-0.100	-3.505	0.001	4.000
7.206	7.255	-0.049	-1.712	0.088	9.000
7.441	7.490	-0.048	-1.681	0.094	17.000
6.568	6.616	-0.048	-1.673	0.095	45.000
5.473	5.517	-0.044	-1.529	0.127	54.000
8.013	8.050	-0.037	-1.296	0.196	25.000
6.272	6.286	-0.014	-0.486	0.627	57.000
8.987	8.993	-0.007	-0.228	0.820	34.000
6.686	6.690	-0.004	-0.125	0.901	22.000
7.083	7.083	0.000	0.004	0.997	6.000
6.939	6.935	0.004	0.134	0.893	31.000

7.981	7.973	0.008	0.269	0.788	24.000
7.555	7.540	0.014	0.496	0.620	Bay State Gas Company
8.750	8.713	0.036	1.273	0.204	5.000 D.T.E. 05-27
7.601	7.565	0.037	1.286	0.199	Attachment RR-DTE-59
8.698	8.655	0.043	1.492	0.136	1.000 Page 5 of 5
8.408	8.361	0.046	1.621	0.106	13.000
7.970	7.916	0.054	1.887	0.060	41.000
9.741	9.652	0.089	3.126	0.002	43.000
6.962	6.867	0.095	3.305	0.001	27.000
7.085	6.968	0.117	4.097	0.000	33.000
7.163	7.031	0.132	4.612	0.000	11.000
8.639	8.488	0.151	5.272	0.000	10.000
7.940	7.770	0.169	5.910	0.000	2.000
8.710	8.532	0.178	6.201	0.000	28.000
8.237	8.029	0.208	7.268	0.000	16.000
8.565	8.351	0.214	6.683	0.000	21.000
7.497	7.226	0.271	9.455	0.000	29.000
7.936	7.639	0.297	10.372	0.000	36.000
8.554	8.108	0.446	15.582	0.000	3.000
					42.000

>>

COMMONWEALTH OF MASSACHUSETTS
DEPARTMENT OF TELECOMMUNICATIONS AND ENERGY

RESPONSE OF BAY STATE GAS COMPANY TO
RECORD REQUESTS FROM THE D.T.E.
D.T.E. 05-27

Date: July 20, 2005

Responsible: Lawrence R. Kaufmann, Consultant (PBR)

RR-DTE-60: Refer to the Company's response to DTE-4-19 and RR-DTE-33. Explain how the assumption that the mean of the random noise in each of the two separate multi-year periods (subsamples) is zero, is consistent with the classical regression model assumptions. Provide a statistical proof of these assumptions. Also, provide a book chapter which deals with this issue, to support the response.

Response: The basic form of PEG's regression model is the following:

$$C_{it} = X_{it}\beta + \eta_{it}$$

where $\eta_{it} = \mu_i + \varepsilon_{it}$

In this model ε_{it} is the random noise component of the error term while μ_i is the firm-specific component, which in our model is equal to $\text{Inefficiency}^i - \text{Inefficiency}^{\text{average}}$. It is standard in such models to assume that $E(\varepsilon_{it}) = 0$ and $E(\mu_i) = 0$, where E is the expectations operator which generates the mean value of a term. One text that deals with this issue, and supports PEG's assumptions, is Greene, W.H., 2000, *Econometric Analysis*, Prentice-Hall: New Jersey; please see p. 568 of Chapter 14 provided in Attachment RR-DTE-60 (a). The econometrics textbook from Arthur Goldberger (*A Course in Econometrics* (1991), Harvard University Press: Cambridge, MA) also supports the assumption that the expected value of the random noise term in the classical regression model is zero; see especially p. 170 of Chapter 15 in Attachment RR=DTE-60 (b).

The assumption that $E(\varepsilon_{it}) = 0$ indicates that the random noise component is expected to equal zero in any given time period t for a given firm i . It follows that, over any multi-year period, the average value of the random noise component will be expected to be zero for a given firm i . Since this result applies to any multi-year period, it holds true for each of the two multi-year periods (subsamples) employed in PEG's analysis.

C H A P T E R

Models for Panel Data

14.1. Introduction

Data sets that combine time series and cross sections are common in economics. For example, the published statistics of the OECD contain numerous series of economic aggregates observed yearly for many countries. Recently constructed **longitudinal** data sets contain observations on thousands of individuals or families, each observed at several points in time. Some empirical studies have analyzed time-series data on several firms, states, or industries simultaneously. These data sets provide a rich source of information about the economy. Modeling in this setting, however, calls for some quite complex stochastic specifications. In this and the next chapter, we will survey the most commonly used techniques for time-series cross-section data analyses. This chapter will describe several techniques that have been applied in single equation models. In Chapter 15, we will consider some different models that also employ time-series cross-section data, but in models that involve sets of equations.

14.2. Panel Data Models

Many recent studies have analyzed **panel**, or longitudinal, data sets. Two very famous ones are the National Longitudinal Survey of Labor Market Experience (NLS) and the Michigan Panel Study of Income Dynamics (PSID). In these data sets, very large cross sections, consisting of thousands of microunits, are followed through time, but the number of periods is often quite small. The PSID, for example, is a study of roughly 6000 families and 15,000 individuals who have been interviewed periodically from 1968 to the present. Another group of intensively studied panel data sets is those from the negative income tax experiments of the early 1970s in which thousands of families were followed for 8 or 13 quarters. Constructing long, evenly spaced time series in this context would be prohibitively expensive, but for the purposes for which these data are typically used, it is unnecessary. Time effects are often viewed as "transitions" or discrete changes of state. They are typically modeled as specific to the period in which they occur and are not carried across periods within a cross-sectional unit.¹ Panel data sets are more oriented toward cross-section analyses; they are wide

¹Theorists have not been prevented from devising autocorrelation models applicable to panel data sets; though. See, for example, Lee (1978). As a practical matter, however, the empirical literature in this field has tended to concentrate on the less intricate models. Time-series modeling of the sort discussed in Chapter 13 is somewhat unusual in the analysis of longitudinal data.

but typically short. Heterogeneity across units is an integral part—indeed, often the central focus—of the analysis.

The analysis of panel or longitudinal data is the subject of one of the most active and innovative bodies of literature in econometrics,² partly because panel data provide such a rich environment for the development of estimation techniques and theoretical results. In more practical terms, however, researchers have been able to use time-series cross-sectional data to examine issues that could not be studied in either cross-sectional or time-series settings alone. Two examples are as follows.

1. In a widely cited study of labor supply, Ben-Porath (1973) observes that at a certain point in time, in a cohort of women, 50 percent may appear to be working. It is ambiguous whether this finding implies that, in this cohort, one-half of the women on average will be working or that the same one-half will be working in every period. These have very different implications for policy and for the interpretation of any statistical results. Cross-sectional data alone will not shed any light on the question.
2. A long-standing problem in the analysis of production functions has been the inability to separate economies of scale and technological change.³ Cross-sectional data provide information only about the former, whereas time-series data muddle the two effects, with no prospect of separation. It is common, for example, to assume constant returns to scale so as to reveal the technical change.⁴ Of course, this practice assumes away the problem. A study by Greene (1983) examines the cost of electric power generation for a large number of firms, each observed in each of several years. The basic model, for the i th firm in year t ,

$$\text{cost}_{it} = C(Y_{it}, \mathbf{p}_{it}, t),$$

where Y is output and \mathbf{p} is a vector of factor prices, provides estimates of the rate of technological change

$$\delta_t = \frac{-d \ln C}{dt}$$

and economies of scale

$$\text{e.s.}_{it} = \frac{1}{\left(\frac{d \ln C_{it}}{d \ln Y_{it}} \right)} - 1.$$

²The panel data literature rivals the received research on unit roots in econometrics in its rate of growth. A compendium of the earliest literature is Maddala (1993). Book-length surveys on the econometrics of panel data include Hsiao (1986), Dielman (1989), Matyas and Sevestre (1996), Raj and Baltagi (1992), and Baltagi (1995). There are also lengthy surveys devoted to specific topics, such as limited dependent variable models and semiparametric methods. An extensive bibliography is given in Baltagi (1995).

³The distinction between these two effects figured prominently in the policy question of whether it was appropriate to break up the AT&T Corporation in the 1980s and, ultimately, to allow competition in the provision of long-distance telephone service.

⁴In a classic study of this issue, Solow (1957) states: "From time series of \dot{Q}/Q , w_K/K , w_L/L and \dot{L}/L or their discrete year-to-year analogues, we could estimate \dot{A}/A and thence $A(t)$ itself. Actually an amusing thing happens here. Nothing has been said so far about returns to scale. But if all factor inputs are classified either as K or L , then the available figures always show w_K and w_L adding up to one. Since we have assumed that factors are paid their marginal products, this amounts to assuming the hypothesis of Euler's theorem. The calculus being what it is, we might just as well assume the conclusion, namely, the F is homogeneous of degree one."

In principle, the methods of Section 13.3 can be applied to longitudinal data sets. In the typical panel, however, there are a large number of cross-sectional units and only a few periods. Thus, the time-series methods discussed there may be somewhat problematic. Recent work has generally concentrated on models better suited to these short and wide data sets. The techniques are focused on cross-sectional variation, or heterogeneity. In this chapter, we shall examine the two most widely used models, then look briefly at some extensions.

EXAMPLE 14.1 Cost Function for Airline Production

The data in Appendix Table A14.1 were used in a study of efficiency in production of airline services in Greene (1997). The airline industry has been a favorite subject of study [e.g., Schmidt and Sickles (1984); Sickles, Good, and Johnson (1986)], partly because of interest in this rapidly changing market in a period of deregulation and partly because of an abundance of large, high-quality data sets collected by the (no longer existent) Civil Aeronautics Board. The original data set consisted of 25 firms observed yearly for 15 years (1970 to 1984), a “balanced panel.” Several of the firms merged during this period and several others experienced strikes, which reduced the number of complete observations substantially. Omitting these and others because of missing data on some of the variables left a group of 10 full observations, from which we have selected six for the examples.

We will examine a simple model for the total cost of production:

$$\log \text{cost}_{it} = \beta_1 + \beta_2 \log \text{output}_{it} + \beta_3 \log \text{fuel price}_{it} + \beta_4 \text{load factor}_{it} + \epsilon_{it}.$$

Output is measured in “revenue passenger miles.” The load factor is a rate of capacity utilization; it is the average rate at which seats on the airline’s planes are filled. More complete models of costs include other factor prices (materials, capital) and, perhaps, a quadratic term in log output to allow for variable economies of scale. We have restricted the cost function to these few variables to provide a straightforward illustration.

Ordinary least squares regression produces the following results. Estimated standard errors are given in parentheses.

$$\begin{aligned} \log \text{cost}_{it} = & 9.5169 + 0.88274 \log \text{output}_{it} + 0.45398 \log \text{fuel price}_{it} \\ & (0.22924) (0.013255) \quad (0.020304) \\ & - 1.62751 \text{load factor}_{it} + \epsilon_{it} \\ & (0.34540) \end{aligned}$$

$$R^2 = 0.98829, s^2 = 0.015528, \mathbf{e'e} = 1.335442$$

The results so far are what one might expect. There are substantial economies of scale; $\text{e.s.}_{it} = (1/0.88274) - 1 = 0.1329$. The fuel price and load factors affect costs in the predictable fashions as well. (Fuel prices differ because of different mixes of types and regional differences in supply characteristics.)

The fundamental advantage of a panel data set over a cross section is that it will allow the researcher far greater flexibility in modeling differences in behavior

across individuals. The basic framework for this discussion is a regression model of the form

$$y_{it} = \alpha_i + \beta'x_{it} + \epsilon_{it}. \quad (14-1)$$

There are K regressors in x_{it} , *not including the constant term*. The **individual effect** is α_i , which is taken to be constant over time t and specific to the individual cross-sectional unit i . As it stands, this model is a classical regression model. If we take the α_i 's to be the same across all units, then ordinary least squares provides consistent and efficient estimates of α and β . There are two basic frameworks used to generalize this model. The **fixed effects** approach takes α_i to be a group specific constant term in the regression model. The **random effects** approach specifies that α_i is a group specific disturbance, similar to ϵ_{it} except that for each group, there is but a single draw that enters the regression identically in each period. We will consider these two approaches in turn.

14.3. Fixed Effects

A common formulation of the model assumes that differences across units can be captured in differences in the constant term.⁵ Thus, in (14-1), each α_i is an unknown parameter to be estimated. Let y_i and X_i be the T observations for the i th unit, and let ϵ_i be associated $T \times 1$ vector of disturbances. Then we may write (14-1) as

$$y_i = i\alpha_i + X_i\beta + \epsilon_i.$$

Collecting these terms gives

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} i & 0 & \cdots & 0 \\ 0 & i & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & i \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$y = [d_1 \ d_2 \ \dots \ d_n \ X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \epsilon, \quad (14-2)$$

where d_i is a dummy variable indicating the i th unit. Let the $nT \times n$ matrix $D = [d_1 \ d_2 \ \dots \ d_n]$. Then, assembling all nT rows gives

$$y = D\alpha + X\beta + \epsilon. \quad (14-3)$$

This model is usually referred to as the **least squares dummy variable (LSDV) model** (although the "least squares" part of the name refers to the technique usually used to estimate it, not to the model as such).

⁵It is also possible to allow the slopes to vary across i , but this method introduces some new methodological issues, as well as considerable complexity in the calculations. A study on the topic is Cornwell and Schmidt (1984). Also, the assumption of a fixed T is only for convenience. The more general case in which T_i varies across units is considered later, in the exercises, and in Greene (1995a).

This model is a classical regression model, so no new results are needed to analyze it. If n is small enough, then the model can be estimated by ordinary least squares with K regressors in \mathbf{X} and n columns in \mathbf{D} , as a multiple regression with $n + K$ parameters. Of course, if n is thousands, as is typical, then this model is likely to exceed the storage capacity of any computer. But, as we found in Chapter 8, there is an easier way to proceed. Using familiar results for a partitioned regression,⁶ we write the OLS estimator of β as

$$\mathbf{b} = [\mathbf{X}'\mathbf{M}_d\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_d\mathbf{y}], \quad (14-4)$$

where

$$\mathbf{M}_d = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'.$$

This amounts to a least squares regression using the transformed data $\mathbf{X}_* = \mathbf{M}_d\mathbf{X}$ and $\mathbf{y}_* = \mathbf{M}_d\mathbf{y}$. The structure of \mathbf{D} is particularly convenient; its columns are orthogonal, so

$$\mathbf{M}_d = \begin{bmatrix} \mathbf{M}^0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ & & \vdots & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}^0 \end{bmatrix}.$$

Each matrix on the diagonal is

$$\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T}\mathbf{ii}'.$$

Premultiplying any $T \times 1$ vector \mathbf{z}_i by \mathbf{M}^0 creates $\mathbf{M}^0\mathbf{z}_i = \mathbf{z}_i - \bar{\mathbf{z}}\mathbf{i}$. (Note that the mean is taken over only the T observations for unit i .) Therefore, the regression of $\mathbf{M}_d\mathbf{y}$ on $\mathbf{M}_d\mathbf{X}$ is equivalent to the regression of $[\mathbf{y}_{it} - \bar{y}_i]$ on $[\mathbf{x}_{it} - \bar{\mathbf{x}}_i]$, where $\bar{\mathbf{x}}_i$ is the $K \times 1$ vector of means of \mathbf{x}_{it} over the T observations. The dummy variable coefficients can be recovered from the other normal equation in the partitioned regression:

$$\mathbf{D}'\mathbf{D}\mathbf{a} + \mathbf{D}'\mathbf{X}\mathbf{b} = \mathbf{D}'\mathbf{y} \quad (14-5)$$

or

$$\mathbf{a} = [\mathbf{D}'\mathbf{D}]^{-1}\mathbf{D}'(\mathbf{y} - \mathbf{X}\mathbf{b}).$$

This implies that for each i ,

$$a_i = \text{the mean residual in the } i\text{th group.} \quad (14-6)$$

Alternatively,

$$a_i = \bar{y}_i - \mathbf{b}'\bar{\mathbf{x}}_i.$$

The appropriate estimator of the covariance matrix for \mathbf{b} is

$$\text{Est. Var}[\mathbf{b}] = s^2[\mathbf{X}'\mathbf{M}_d\mathbf{X}]^{-1}, \quad (14-7)$$

⁶See Section 6.4.3.

which uses the usual second moment matrix with \mathbf{x} 's, now expressed as deviations from their respective unit means. The disturbance variance estimator is

$$s^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it} - a_i - \mathbf{x}'_{it} \mathbf{b})^2}{nT - n - K}. \quad (14-8)$$

The it th residual is

$$\begin{aligned} e_{it} &= y_{it} - a_i - \mathbf{x}'_{it} \mathbf{b} \\ &= y_{it} - (\bar{y}_i - \bar{\mathbf{x}}'_i \mathbf{b}) - \mathbf{x}'_{it} \mathbf{b} \\ &= (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \mathbf{b}. \end{aligned}$$

Thus, the numerator in s^2 is exactly the sum of squared residuals from the regression in (14-4). But most computer programs will use $nT - K$ for the denominator in computing s^2 , so a correction will be necessary. For the individual effects,

$$\text{Var}[a_i] = \frac{\sigma^2}{T} + \bar{\mathbf{x}}'_i \text{Var}[\mathbf{b}] \bar{\mathbf{x}}_i,$$

so a simple estimator based on s^2 can be computed.

14.3.1. TESTING THE SIGNIFICANCE OF THE GROUP EFFECTS

The usual t ratio for a_i implies a test of the hypothesis that a_i equals zero. This hypothesis, however, is typically not useful for testing in a regression context. If we are interested in differences across groups, then we can test the hypothesis that the constant terms are all equal with an F test. Under the null hypothesis, the efficient estimator is pooled least squares. The F ratio used for the test is

$$F(n-1, nT-n-K) = \frac{(R_u^2 - R_p^2)/(n-1)}{(1 - R_u^2)/(nT-n-K)}, \quad (14-9)$$

where u indicates the unrestricted model and p indicates the pooled or restricted model with only a single overall constant term. (The sums of squared residuals may be used instead if that is more convenient.) It may be more convenient to estimate the model with an overall constant and $n-1$ dummy variables instead. The other results will be unchanged, and rather than estimate a_i , each dummy variable coefficient will be an estimate of $a_i - a_1$. The F test that the coefficients on the $n-1$ dummy variables are zero is identical to the one above. It is important to keep in mind that although the statistical results are the same, the interpretation of the dummy variable coefficients in the two formulations is different.⁷

14.3.2. THE WITHIN AND BETWEEN GROUPS ESTIMATORS

We could formulate a pooled regression model in three ways. First, the original formulation is

$$y_{it} = \alpha + \beta' \mathbf{x}_{it} + \epsilon_{it}. \quad (14-10a)$$

⁷For a discussion of the differences, see Suits (1984).

In terms of deviations from the group means,

$$y_{it} - \bar{y}_{i.} = \beta'(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.}) + \epsilon_{it} - \bar{\epsilon}_{i.}, \quad (14-10b)$$

whereas in terms of the group means,

$$\bar{y}_{i.} = \alpha + \beta' \bar{\mathbf{x}}_{i.} + \bar{\epsilon}_{i.}. \quad (14-10c)$$

All three are classical regression models, and in principle, all three could be estimated, at least consistently if not efficiently, by ordinary least squares. [Note that (14-10c) involves only n observations, the group means.] Consider then the matrices of sums of squares and cross products that would be used in each case, where we focus only on estimation of β . In (14-10a), the moments would be about the overall means, \bar{y} and $\bar{\mathbf{x}}$, and we would use the total sums of squares and cross products,

$$\mathbf{S}'_{xx} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})'$$

and

$$\mathbf{S}'_{xy} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(y_{it} - \bar{y}).$$

(The use of the superscript t indicates "total" and is unrelated to the time subscript.) For (14-10b), since the data are in deviations already, the means of $(y_{it} - \bar{y}_{i.})$ and $(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})$ are zero. The moment matrices are **within-groups** (i.e., deviations from group means) sums of squares and cross products,

$$\mathbf{S}^w_{xx} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})'$$

and

$$\mathbf{S}^w_{xy} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(y_{it} - \bar{y}_{i.}).$$

Finally, for (14-10c), the mean of group means is the overall mean. The moment matrices are the **between-groups** sums of squares and cross products,

$$\mathbf{S}^b_{xx} = \sum_{i=1}^n T(\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}})'$$

and

$$\mathbf{S}^b_{xy} = \sum_{i=1}^n T(\bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}})(\bar{y}_{i.} - \bar{y}).$$

It is easy to verify that

$$\mathbf{S}'_{xx} = \mathbf{S}^w_{xx} + \mathbf{S}^b_{xx}$$

and

$$\mathbf{S}'_{xy} = \mathbf{S}^w_{xy} + \mathbf{S}^b_{xy}.$$

There are, therefore, three possible least squares estimators of β corresponding to the decomposition. The least squares estimator is

$$\mathbf{b}^l = [\mathbf{S}_{xx}^l]^{-1} \mathbf{S}_{xy}^l = [\mathbf{S}_{xx}^w + \mathbf{S}_{xx}^b]^{-1} [\mathbf{S}_{xy}^w + \mathbf{S}_{xy}^b]. \quad (14-11)$$

The **within-groups** estimator is

$$\mathbf{b}^w = [\mathbf{S}_{xx}^w]^{-1} \mathbf{S}_{xy}^w. \quad (14-12)$$

This is the LSDV estimator computed earlier. [See (14-4).] An alternative estimator would be the **between-groups** estimator,

$$\mathbf{b}^b = [\mathbf{S}_{xx}^b]^{-1} \mathbf{S}_{xy}^b \quad (14-13)$$

(sometimes called the **group means** estimator). This least squares estimator of (14-10c) is based on the n sets of groups means. From the preceding expressions (and familiar previous results),

$$\mathbf{S}_{xy}^w = \mathbf{S}_{xx}^w \mathbf{b}^w$$

and

$$\mathbf{S}_{xy}^b = \mathbf{S}_{xx}^b \mathbf{b}^b.$$

inserting these in (14-11), we see that the OLS estimator is a matrix weighted average of the within- and between-groups estimators:

$$\mathbf{b}^l = \mathbf{F}^w \mathbf{b}^w + \mathbf{F}^b \mathbf{b}^b, \quad (14-14)$$

where

$$\mathbf{F}^w = [\mathbf{S}_{xx}^w + \mathbf{S}_{xx}^b]^{-1} \mathbf{S}_{xx}^w = \mathbf{I} - \mathbf{F}^b.$$

14.3.3. FIXED TIME AND GROUP EFFECTS

The least squares dummy variable approach can be extended to include a time-specific effect as well. One way to formulate the extended model is simply to add the time effect, as in

$$y_{it} = \alpha_i + \gamma_t + \beta' \mathbf{x}_{it} + \epsilon_{it}. \quad (14-15)$$

This model is obtained from the preceding one by the inclusion of an additional $T - 1$ dummy variables. One of the time effects must be dropped to avoid perfect collinearity. If the number of variables is too large to handle by ordinary regression, then this model can also be estimated by using the partitioned regression.⁸ There is an asymmetry in this formulation, however, since each of the group effects is a group-

⁸The matrix algebra and the theoretical development of two-way effects in panel data models are quite complex. See, for example, Baltagi (1995). Fortunately, the practical application is much simpler. The number of periods analyzed in most panel data sets is rarely more than a handful. Since modern computer programs, even those written strictly for microcomputers, uniformly allow dozens (or even hundreds) of regressors, almost any application involving a second fixed effect can be handled just by literally including the second effect as a set of actual dummy variables.

specific intercept, whereas the time effects are **contrasts**, that is, comparisons to a base period (the one that is excluded). A symmetric form of the model is

$$y_{it} = \mu + \alpha_i + \gamma_t + \beta'x_{it} + \epsilon_{it}, \quad (14-15')$$

where a full n and T effects are included, but the restrictions

$$\sum_i \alpha_i = \sum_t \gamma_t = 0$$

are imposed. Least squares estimates of the slopes are obtained by regression of

$$y_{*it} = y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{\bar{y}} \quad (14-16)$$

on

$$x_{*it} = x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{\bar{x}},$$

where

$$\bar{y}_{i.} = \frac{1}{n} \sum_{t=1}^T y_{it},$$

$$\bar{\bar{y}} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{it},$$

and likewise for $\bar{x}_{i.}$ and $\bar{\bar{x}}$. The overall constant and the dummy variable coefficients can be recovered from the normal equations as

$$m = \bar{\bar{y}} - \mathbf{b}'\bar{\bar{x}},$$

$$a_i = (\bar{y}_{i.} - \bar{\bar{y}}) - \mathbf{b}'(\bar{x}_{i.} - \bar{\bar{x}}), \quad (14-17)$$

$$c_t = (\bar{y}_{.t} - \bar{\bar{y}}) - \mathbf{b}'(\bar{x}_{.t} - \bar{\bar{x}}).$$

The estimated covariance matrix for \mathbf{b} is computed using the sums of squares and cross products of x_{*it} and s^2 computed as usual, $\mathbf{e}'\mathbf{e}/[nT - (n - 1) - (T - 1) - K - 1]$. If one of n or T is small and the other is large, then it will usually be simpler just to treat the smaller set as an ordinary set of variables and apply the previous results to the one-way fixed effects model defined by the larger set. Although more general, this model is rarely used in practice. There are two reasons. First, the cost in terms of degrees of freedom is often not justified. Second, in those instances in which a model of the timewise evolution of the disturbance is desired, a more general model than the dummy variable formulation is usually used.

EXAMPLE 14.2 Fixed Effects Regressions

Table 14.1 contains the estimated cost equations with individual firm effects, specific period effects, and both firm and period effects. For comparison, the least squares and group means results are given also. The F statistic for testing the joint significance of the firm effects is

$$F[5, 81] = \frac{(0.99743 - 0.98829)/5}{(1 - 0.99743)/81} = 57.614.$$

The critical value from the F table is 2.327, so the evidence is strongly in favor of a firm specific effect in the data. The same computation for the time effects, in the ab-

TABLE 14.1 Cost Equations with Firm and Period Effects

Specification	Parameter Estimates							
	β_1	β_2	β_3	β_4	R^2	s^2		
No effects	9.517 (0.22924)	0.88274 (0.013255)	0.45398 (0.020304)	-1.6275 (0.34530)	0.98829	0.015528		
Group means	85.812 (56.482)	0.78246 (0.10877)	-5.5240 (4.47870)	-1.7510 (2.74308)	0.99364	0.015837		
Firm effects		0.91930 (0.029890)	0.41749 (0.015199)	-1.07040 (0.20169)	0.99743	0.0036125		
$a_1 \dots a_6$	9.706	9.665	9.497	9.891	9.730	9.793		
Time effects		0.86773 (0.015408)	-0.48447 (0.36411)	-1.95442 (0.44238)	0.99046	0.015114		
$c_1 \dots c_8$	20.496	20.578	20.656	20.741	21.200	21.411	21.503	21.654
$c_9 \dots c_{15}$	21.829	22.114	22.465	22.651	22.616	22.552	22.537	
Firm and time effects	12.667 (2.0811)	0.81725 (0.031851)	0.16861 (0.16347)	-0.88281 (0.26174)	0.99845	0.0026395		
$a_1 \dots a_6$	0.12833	0.06549	-0.18947	0.13425	-0.09265	-0.04595		
$c_1 \dots c_8$	-0.37402	-0.31932	-0.27669	-0.22304	-0.15393	0.10809	-0.07686	-0.02073
$c_9 \dots c_{15}$	0.04722	0.09173	0.20731	0.28547	0.30138	0.30047	0.31911	

sence of the firm effects produces an $F[14, 72]$ statistic of 1.1685, which is considerably less than the 95 percent critical value of 1.832. Thus, on this basis, there does not appear to be a significant cost difference across the different periods that is not accounted for by the fuel price variable, output, and load factors. There is a distinctive pattern to the time effects, which we will examine more closely later. In the presence of the firm effects, the $F[14, 72]$ ratio for the joint significance of the period effects is 3.133, which is larger than the table value of 1.842.

14.3.4. UNBALANCED PANELS AND FIXED EFFECTS

Missing data are very common in panel data sets. For this reason, or perhaps just because of the way the data were recorded, panels in which the group sizes differ across groups are not unusual. These panels are called **unbalanced panels**. The preceding analysis assumed equal group sizes and relied on the assumption at several points. A modification to allow unequal group sizes is quite simple.⁹ First, the full sample size is $\sum_{i=1}^n T_i$ instead of nT , which calls for minor modifications in the computations of s^2 .

⁹Since most modern econometrics computer packages fully automate the computation, this discussion is presented in the interests of removing the veil over what sometimes appears to be a fairly arcane set of calculations.

$\text{Var}[\mathbf{b}]$, $\text{Var}[a_i]$, and the F statistic. Second, group means must be based on T_i , which varies across groups. The overall means for the regressors are simply

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^n \sum_{t=1}^{T_i} \mathbf{x}_{it}}{\sum_{i=1}^n T_i} = \frac{\sum_{i=1}^n T_i \bar{\mathbf{x}}_i}{\sum_{i=1}^n T_i} = \sum_{i=1}^n w_i \bar{\mathbf{x}}_i,$$

where $w_i = T_i / (\sum_{i=1}^n T_i)$. If the group sizes are equal, then $w_i = 1/n$. The moment matrix shown in (14-4),

$$\mathbf{S}_{xx}^w = \mathbf{X}' \mathbf{M}_d \mathbf{X},$$

is a sum of matrices of sums of squares and cross products, summed across the groups,

$$\sum_{i=1}^n \mathbf{X}_i' \mathbf{M}_i^0 \mathbf{X}_i = \sum_{i=1}^n \left(\sum_{t=1}^{T_i} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right).$$

This matrix is called the **within-groups** sum of squares. The other moments, \mathbf{S}_{xy}^w and \mathbf{S}_{yy}^w , are computed likewise. No other changes are necessary for the one factor LSDV estimator. The two-way model can be handled likewise, although with unequal group sizes in both directions, the algebra becomes fairly cumbersome. Once again, however, the practice is much simpler than the theory. The easiest approach for unbalanced panels is just to create the full set of T dummy variables using as T the union of the dates represented in the full data set. One (presumably the last) is dropped, so we revert back to (14-15). Then, within each group, any of the T periods represented is accounted for by using one of the dummy variables. Least squares using the LSDV approach for the group effects will then automatically take care of the messy accounting details.

14.4. Random Effects

The fixed effects model is a reasonable approach when we can be confident that the differences between units can be viewed as parametric shifts of the regression function. This model might be viewed as applying only to the cross-sectional units in the study, not to additional ones outside the sample. For example, an intercountry comparison may well include the full set of countries for which it is reasonable to assume that the model is constant. In other settings, it might be more appropriate to view individual specific constant terms as randomly distributed across cross-sectional units. This view would be appropriate if we believed that sampled cross-sectional units were drawn from a large population. It would certainly be the case for the longitudinal data sets listed in the introduction to this chapter.¹⁰

Consider, then, a reformulation of the model

$$y_{it} = \alpha + \beta' \mathbf{x}_{it} + u_i + \epsilon_{it}, \quad (14-18)$$

¹⁰This distinction is not hard and fast; it is purely heuristic. We shall return to this issue later.

where there are K regressors in addition to the constant term. The component u_i is the random disturbance characterizing the i th observation and is constant through time. In the analysis of families, we can view them as the collection of factors not in the regression that are specific to that family. We assume further that

$$\begin{aligned} E[\epsilon_{it}] &= E[u_i] = 0, \\ E[\epsilon_{it}^2] &= \sigma_\epsilon^2, \\ E[u_i^2] &= \sigma_u^2, \\ E[\epsilon_{it}u_j] &= 0 \quad \text{for all } i, t, \text{ and } j, \\ E[\epsilon_{it}\epsilon_{js}] &= 0 \quad \text{if } t \neq s \text{ or } i \neq j, \\ E[u_i u_j] &= 0 \quad \text{if } i \neq j. \end{aligned} \quad (14-19)$$

As before, it is useful to view the formulation of the model in blocks of T observations for observations i , y_i , X_i , u_i , and ϵ_i . For these T observations, let

$$w_{it} = \epsilon_{it} + u_i$$

and

$$w_i = [w_{i1}, w_{i2}, \dots, w_{iT}]'$$

In view of this form of w_{it} , we have what is often called an "error components model." Then, for this model,

$$\begin{aligned} E[w_{it}^2] &= \sigma_\epsilon^2 + \sigma_u^2, \\ E[w_{it}w_{is}] &= \sigma_u^2, \quad t \neq s. \end{aligned}$$

For the T observations for unit i , let $\Omega = E[w_i w_i']$. Then

$$\Omega = \begin{bmatrix} \sigma_\epsilon^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\epsilon^2 + \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_\epsilon^2 + \sigma_u^2 \end{bmatrix} = \sigma_\epsilon^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}_T', \quad (14-20)$$

where \mathbf{i} is a $T \times 1$ column vector of 1s. Since observations i and j are independent, the disturbance covariance matrix for the full nT observations is

$$\mathbf{V} = \begin{bmatrix} \Omega & 0 & 0 & \dots & 0 \\ 0 & \Omega & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Omega \end{bmatrix} = \Omega \otimes \mathbf{I}_n. \quad (14-21)$$

This matrix has a particularly simple structure.

14.4.1. GENERALIZED LEAST SQUARES

For generalized least squares, we require $\mathbf{V}^{-1/2} = \mathbf{I} \otimes \Omega^{-1/2}$. Therefore, we need only find $\Omega^{-1/2}$, which is

$$\Omega^{-1/2} = \frac{1}{\sigma_\epsilon} \left[\mathbf{I} - \frac{\theta}{T} \mathbf{i} \mathbf{i}' \right],$$

where

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{T\sigma_u^2 + \sigma_\epsilon^2}}.$$

The transformation of y_i and X_i for GLS is therefore

$$\Omega^{-1/2}y_i = \frac{1}{\sigma_\epsilon} \begin{bmatrix} y_{i1} - \theta\bar{y}_i \\ y_{i2} - \theta\bar{y}_i \\ \vdots \\ y_{iT} - \theta\bar{y}_i \end{bmatrix}, \quad (14-22)$$

and likewise for the rows of X_i .¹¹ For the data set as a whole, generalized least squares is computed by the regression of these partial deviations of y_{it} on the same transformations of x_{it} . Note the similarity of this procedure to the computation in the LSDV model, which has $\theta = 1$. (One could interpret θ as the effect that would remain if σ_ϵ were zero, because the only effect would then be u_i . In this case, the fixed and random effects models would be indistinguishable, so this result makes sense.)

It can be shown that the GLS estimator is, like the OLS estimator, a matrix weighted average of the within- and between-units estimators:

$$\hat{\beta} = \hat{F}^w b^w + (\mathbf{I} - \hat{F}^w) b^b,^{12} \quad (14-23)$$

where

$$\hat{F}^w = [S_{xx}^w + \lambda S_{xx}^b]^{-1} S_{xx}^w, \\ \lambda = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T\sigma_u^2} = (1 - \theta)^2.$$

To the extent that λ differs from one, we see that the inefficiency of least squares will follow from an inefficient weighting of the two least squares estimators. Compared with generalized least squares, ordinary least squares places too much weight on the between-units variation. It includes it all in the variation in X , rather than apportioning some of it to random variation across groups attributable to the variation in u_i across units.

There are some polar cases to consider. If λ equals 1, then generalized least squares is ordinary least squares. This situation would occur if σ_u^2 were zero, in which case a classical regression model would apply. If λ equals zero, then the estimator is the dummy variable estimator we used in the fixed effects setting. There are two possibilities. If σ_ϵ^2 were zero, then all variation across units would be due to the different u_i 's, which, because they are constant across time, would be equivalent to the dummy variables we used in the fixed-effects model. The question of whether they were fixed or random would then become moot. They are the only source of variation across units once the regression is accounted for. The other case is $T \rightarrow \infty$. We can view it

¹¹This transformation is a special case of the more general treatment in Nerlove (1971b).

¹²An alternative form of this expression, in which the weighing matrices are proportional to the covariance matrices of the two estimators, is given by Judge et al. (1985).

this way: If $T \rightarrow \infty$, then the unobserved u_i becomes observable. Take the T observations for the i th unit. Our estimator of $[\alpha, \beta]$ is consistent in the dimensions T or n . Therefore,

$$y_{it} - \alpha - \beta'x_{it} = u_i + \epsilon_{it}$$

is observable. The individual means will provide

$$\bar{y}_{i.} - \alpha - \beta'\bar{x}_{i.} = u_i + \bar{\epsilon}_{i.}$$

But $\bar{\epsilon}_{i.}$ converges to zero, which reveals u_i to us. Therefore, if T goes to infinity, u_i becomes the d_i we used earlier. (That it is not 1 is immaterial; it is nonzero only for unit i .)

14.4.2. FEASIBLE GENERALIZED LEAST SQUARES WHEN Ω IS UNKNOWN

If the variance components are known, generalized least squares can be computed without much difficulty. Of course, it is not likely, so as usual, we must first estimate the disturbance variances and then use an FGLS procedure. A heuristic approach to estimation of the components is as follows:

$$y_{it} = \alpha + \beta'x_{it} + \epsilon_{it} + u_i$$

and

$$\bar{y}_{i.} = \alpha + \beta'\bar{x}_{i.} + \bar{\epsilon}_{i.} + u_i. \quad (14-24)$$

Therefore, taking deviations from the group means removes the heterogeneity:

$$y_{it} - \bar{y}_{i.} = \beta'[x_{it} - \bar{x}_{i.}] + [\epsilon_{it} - \bar{\epsilon}_{i.}]. \quad (14-25)$$

Since

$$E\left[\sum_{t=1}^T (\epsilon_{it} - \bar{\epsilon}_{i.})^2\right] = (T-1)\sigma_\epsilon^2,$$

if β were observed, then an unbiased estimator of σ_ϵ^2 based on T observations in group i would be

$$\hat{\sigma}_\epsilon^2(i) = \frac{\sum_{t=1}^T (\epsilon_{it} - \bar{\epsilon}_{i.})^2}{T-1}. \quad (14-26)$$

Since β must be estimated—we use the LSDV estimator in (14-4)—we make the usual degrees of freedom correction and use

$$s_c^2(i) = \frac{\sum_{t=1}^T (\epsilon_{it} - \bar{\epsilon}_{i.})^2}{T-K-1}. \quad (14-27)$$

We have n such estimators, so we average them to obtain

$$\begin{aligned}\bar{s}_e^2 &= \frac{1}{n} \sum_{i=1}^n s_e^2(i) \\ &= \frac{1}{n} \sum_{i=1}^n \left[\frac{\sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{T - K - 1} \right] \\ &= \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{nT - nK - n}.\end{aligned}\tag{14-28}$$

The degrees of freedom correction in \bar{s}_e^2 is excessive because it assumes that α and β are reestimated for each i . The estimated parameters are the n means \bar{y}_i , and the K slopes. Therefore, we propose the unbiased estimator¹³

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{nT - n - K}.\tag{14-29}$$

This is the variance estimator in the LSDV model in (14-8), appropriately corrected for degrees of freedom. The n means,

$$\begin{aligned}\epsilon_{**i} &= \bar{y}_i - \alpha - \beta' \bar{x}_i \\ &= \bar{e}_i + u_{i*},\end{aligned}\tag{14-30}$$

are independent and have variance

$$\text{Var}[\epsilon_{**i}] = \sigma_{**}^2 = \frac{\sigma_e^2}{T} + \sigma_u^2.$$

By incorporating the degrees of freedom correction for the estimate of β in the group means least squares regression of (14-24), we can use

$$\hat{\sigma}_{**}^2 = \frac{\mathbf{e}_{**}' \mathbf{e}_{**}}{n - K} = m_{**}\tag{14-31}$$

as an unbiased estimator of $\sigma_e^2/T + \sigma_u^2$. This suggests the estimator

$$\hat{\sigma}_u^2 = \hat{\sigma}_{**}^2 - \frac{\hat{\sigma}_e^2}{T}.\tag{14-32}$$

The estimator in (14-32) is unbiased but could be negative in a finite sample. Alternative estimators have been proposed.¹⁴ Since we only require a consistent estimator of σ_u^2 , any consistent estimator of β could be used in (14-31), including the original pooled OLS estimator. Such a finding might cast some doubt on the

¹³A formal proof of this proposition may be found in Maddala (1971) or in Judge et al. (1985, p. 551).

¹⁴See, for example, Wallace and Hussain (1969), Maddala (1971), Fuller and Battese (1974), and Amemiya (1971).

appropriateness of the model, however, and before proceeding in this fashion, one might do well to reconsider the random effects specification.

There is a remaining complication. If there are any regressors that do not vary within the groups, the LSDV estimator cannot be computed. For example, in a model of family income or labor supply, one of the regressors might be a dummy variable for location, family structure, or living arrangement. Any of these could be perfectly collinear with the fixed effect for that family, which would prevent computation of the LSDV estimator. In this case, it is still possible to estimate the random effects variance components. Once again, let $[a, b]$ be any consistent estimator of $[\alpha, \beta]$, such as the ordinary least squares estimator. Then, using all nT residuals, $m_{ee} = e'e/(nT)$ has

$$\text{plim} \frac{e'e}{nT} = \sigma_e^2 + \sigma_u^2.$$

Now, using the n group means, (14-31) is still usable for estimation. This result produces two moment equations in the two unknown variance terms,

$$\begin{aligned} m_{**} &= \frac{\sigma_e^2}{T} + \sigma_u^2, \\ m_{ee} &= \sigma_e^2 + \sigma_u^2, \end{aligned}$$

which have solutions

$$\begin{aligned} \hat{\sigma}_e^2 &= \frac{T}{T-1} (m_{ee} - m_{**}) \\ \hat{\sigma}_u^2 &= \frac{T}{T-1} m_{**} - \frac{1}{T-1} m_{ee} \\ &= \omega m_{**} + (1 - \omega) m_{ee}, \end{aligned}$$

where $\omega > 1$. As before, this estimator can produce a negative estimate of σ_u^2 that, once again, calls the specification of the model into question.

14.4.3. TESTING FOR RANDOM EFFECTS

Breusch and Pagan (1980) have devised a Lagrange multiplier test for the random effects model based on the OLS residuals. For

$$\begin{aligned} H_0: \sigma_u^2 &= 0 \quad (\text{or } \text{Corr}[w_{it}, w_{is}] = 0), \\ H_1: \sigma_u^2 &\neq 0, \end{aligned}$$

the test statistic is

$$\begin{aligned} \text{LM} &= \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n \left[\sum_{t=1}^T e_{it} \right]^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \\ &= \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n (T \bar{e}_i)^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2. \end{aligned} \tag{14-33}$$

Under the null hypothesis, LM is distributed as chi-squared with one degree of freedom. A useful shortcut for computing LM is as follows. Let \mathbf{D} be the matrix of dummy variables defined in (14-2), and let \mathbf{e} be the OLS residual vector. Then

$$LM = \frac{nT}{2(T-1)} \left[\frac{\mathbf{e}'\mathbf{D}\mathbf{D}'\mathbf{e}}{\mathbf{e}'\mathbf{e}} - 1 \right]^2. \quad (14-34)$$

EXAMPLE 14.3 Testing for Random Effects

The least squares estimates for the cost equation were given in Example 14.1. The firm specific means of the least squares residuals are

$$\bar{\mathbf{e}} = [0.068869, -0.013878, -0.19422, 0.15273, -0.021583, 0.0080906]'$$

The total sum of squared residuals for the least squares regression is $\mathbf{e}'\mathbf{e} = 1.33544$, so

$$LM = \frac{nT}{2(T-1)} \left[\frac{T^2 \bar{\mathbf{e}}'\bar{\mathbf{e}}}{\mathbf{e}'\mathbf{e}} - 1 \right]^2 = 334.85.$$

Based on the least squares residuals, we obtain a Lagrange multiplier test statistic of 334.85, which far exceeds the 95 percent critical value for chi-squared with one degree of freedom, 3.84. At this point, we conclude that the classical regression model with a single constant term is inappropriate for these data. The result of the test is to reject the null hypothesis in favor of the random effects model. But, it is best to reserve judgment on that, because there is another competing specification that might induce these same results, the fixed effects model. We will examine this possibility in the subsequent examples.

With the variance estimators in hand, FGLS can be used to estimate the parameters of the model. All our earlier results for FGLS estimators apply here. It would also be possible to obtain the maximum likelihood estimator.¹⁵ The likelihood function is complicated, but as we have seen repeatedly, the MLE of β will be GLS based on the maximum likelihood estimators of the variance components. It can be shown that the MLEs of σ_e^2 and σ_u^2 are the unbiased estimators shown earlier, *without* their degrees of freedom corrections.¹⁶ This model satisfies the requirements for the Oberhofer-Kmenta algorithm, so we could also use the iterated FGLS procedure to obtain the MLEs if desired. The initial consistent estimates could be based on least squares residuals. Still other estimators have been proposed. None will have better asymptotic properties than the MLE or FGLS estimators, but they may outperform them in a finite sample.¹⁷

EXAMPLE 14.4 Random Effects Models

To compute the FGLS estimator, we require estimates of the variance components. The unbiased estimator of σ_e^2 is the residual variance estimator in the within-units (LSDV) regression. Thus,

$$\hat{\sigma}_e^2 = \frac{0.2926116}{90 - 9} = 0.0036125$$

¹⁵See Hsiao (1986).

¹⁶See Berzeg (1979).

¹⁷See Maddala and Mount (1973).

The group means regression results were given in Table 14.1. The residual squares in the group means regression is 0.3167626, so

$$\frac{\sigma_e^2}{T} + \sigma_u^2 = \frac{0.03167276}{6 - 4} = 0.015838.$$

Therefore,

$$\hat{\sigma}_u^2 = 0.015838 - \frac{0.0036125}{15} = 0.0155973.$$

For purposes of FGLS,

$$\hat{\theta} = 1 - \left[\frac{0.0036125}{15(0.015838)} \right]^{1/2} = 0.87669.^{18}$$

TABLE 14.2 Random and Fixed Effects Estimates

Specification	Parameter Estimates					
	β_1	β_2	β_3	β_4	R^2	s^2
No effects	9.517 (0.22924)	0.88274 (0.013255)	0.45398 (0.020304)	-1.6275 (0.34530)	0.98829	0.015528
Firm effects	Fixed effects					
		0.91930 (0.029890)	0.41749 (0.015199)	-1.0704 (0.20169)	0.99743	0.0036125
	White(1)	(0.019106)	(0.013533)	(0.21662)		
	White(2)	(0.027977)	(0.013802)	(0.20373)		
	Fixed effects with autocorrelation $\hat{\rho} = 0.5162$					
		0.92804 (0.033112)	0.39197 (0.016911)	-1.21932 (0.20262)		0.001967 $s^2/(1 - \hat{\rho}^2) = 0.002681$
	Random effects					
	9.627 (0.20986)	0.90668 (0.02559)	0.42278 (0.01400)	-1.0645 (0.2000)	$\hat{\sigma}_u^2 = 0.0155963$ $\hat{\sigma}_e^2 = 0.0036125$	
	Random effects with autocorrelation $\hat{\rho} = 0.5162$					
	10.071 (0.2597)	0.91661 (0.029150)	0.39671 (0.01604)	-1.2103 (0.20082)	$\hat{\sigma}_u^2 = 0.026648$ $\hat{\sigma}_e^2 = 0.0074974$	
Firm and time effects	Fixed effects					
	12.667 (2.0811)	0.81725 (0.031851)	0.16861 (0.16347)	-0.88281 (0.26174)	0.99845	0.0026395
	Random effects					
	9.599 (0.19122)	0.90237 (0.023066)	0.42418 (0.01264)	-1.0531 (0.1779)	$\hat{\sigma}_u^2 = 0.015662$ $\hat{\sigma}_e^2 = 0.0026395$ $\hat{\sigma}_v^2 = 0.000068322$	

¹⁸The estimate of θ is a function of two moments of the residuals. The constant in the regression is also a sample moment. Therefore, the results of Section 4.7.2 could be used to estimate the standard error of $a/(1 - \theta)$.

The FGLS estimates for this random effects model are shown in Table 14.2, with the fixed effects estimates. The estimated within-groups variance is larger than the between-groups variance by a factor of five. Thus, by these estimates, over 80 percent of the disturbance variation is explained by variation within the groups, with only the small remainder explained by variation across groups.

None of the desirable properties of the estimators in the random effects model relies on T going to infinity.¹⁹ Indeed, T is likely to be quite small. The maximum likelihood estimator of σ_e^2 is exactly equal to an average of n estimators, each based on the T observations for unit i . [See (14-28).] Each component in this average is, in principle, consistent. That is, its variance is of order $1/T$ or smaller. Since T is small, this variance may be relatively large. Each term provides some information about the parameter, however. The average over the n cross-sectional units has a variance of order $1/(nT)$, which will go to zero if n increases, even if we regard T as fixed. The conclusion to draw is that nothing in this treatment relies on T growing large. Although it can be shown that some consistency results will follow for T increasing, the typical panel data set is based on data sets for which it does not make sense to assume that T increases without bound or, in some cases, at all.²⁰ As a general proposition, it is necessary to take some care in devising estimators whose properties hinge on whether T is large or not. The widely used conventional ones we have discussed here do not, but we have not exhausted the possibilities.

The LSDV model *does* rely on T increasing for consistency. To see, we use the partitioned regression. The slopes are

$$\mathbf{b} = [\mathbf{X}'\mathbf{M}_d\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_d\mathbf{y}].$$

Since \mathbf{X} is $nT \times K$, as long as the inverted moment matrix converges to a zero matrix, \mathbf{b} is consistent as long as either n or T increases without bound. But the dummy variable coefficients are

$$\begin{aligned} a_i &= \bar{y}_i - \mathbf{b}'\bar{\mathbf{x}}_i \\ &= \frac{1}{T} \sum_{t=1}^T e_{it}. \end{aligned}$$

We have already seen that \mathbf{b} is consistent. Suppose, for the present, that $\bar{\mathbf{x}}_i = 0$. Then $\text{Var}[a_i] = \text{Var}[y_{it}]/T$. Therefore, unless $T \rightarrow \infty$, the estimators of the unit-specific effects are not consistent. (They are, however, best linear unbiased.) This inconsistency is worth bearing in mind when analyzing data sets for which T is fixed and there is no intention to replicate the study and no logical argument that would justify the claim that it could have been replicated in principle.

The random effects model was developed by Balestra and Nerlove (1966). Their formulation included a time-specific component as well as the individual effect:

$$y_{it} = \alpha + \beta'x_{it} + \epsilon_{it} + u_i + v_t.$$

¹⁹See Nickell (1981).

²⁰In this connection, Chamberlain (1983) has provided some innovative treatments of panel data that, in fact, take T as given in the model and that base consistency results solely on n increasing. Some additional results for dynamic models are given by Bhargava and Sargan (1983).

The extended formulation is rather complicated analytically. In Balestra and Nerlove's study, it was made even more so by the presence of a lagged dependent variable that causes all the problems discussed earlier in our discussion of autocorrelation. A full set of results for this extended model, including a method for handling the lagged dependent variable, has been developed.²¹ The full model is rarely used, however. Most studies limit the model to the individual effects and, if needed, model the time effects in some other fashion.²²

14.4.4. HAUSMAN'S TEST FOR FIXED OR RANDOM EFFECTS

At various points, we have made the distinction between fixed and random effects models. An inevitable question is, Which should be used? It has been suggested that the distinction between fixed and random effects models is an erroneous interpretation. Mundlak (1978) argues that we should always treat the individual effects as random. The fixed effects model is simply analyzed conditionally on the effects present in the observed sample. One can argue that certain institutional factors or characteristics of the data argue for one or the other, but unfortunately, this approach does not always provide much guidance. From a purely practical standpoint, the dummy variable approach is costly in terms of degrees of freedom lost, and in a wide, longitudinal data set, the random effects model has some intuitive appeal. On the other hand, the fixed effects approach has one considerable virtue. There is no justification for treating the individual effects as uncorrelated with the other regressors, as is assumed in the random effects model. The random effects treatment, therefore, may suffer from the inconsistency due to omitted variables.²³

It is possible to test for orthogonality of the random effects and the regressors. The specification test devised by Hausman (1978)²⁴ has the same format as that for the errors in variables model discussed in Section 9.5.4. It is based on the idea that under the hypothesis of no correlation, both OLS in the LSDV model and GLS are consistent, but OLS is inefficient,²⁵ whereas under the alternative, OLS is consistent, but GLS is not. Therefore, under the null hypothesis, the two estimates should not differ systematically, and a test can be based on the difference. The other essential ingredient for the test is the covariance matrix of the difference vector, $[\mathbf{b} - \hat{\beta}]$:

$$\text{Var}[\mathbf{b} - \hat{\beta}] = \text{Var}[\mathbf{b}] + \text{Var}[\hat{\beta}] - \text{Cov}[\mathbf{b}, \hat{\beta}] - \text{Cov}[\mathbf{b}, \hat{\beta}]'. \quad (14-35)$$

Hausman's essential result is that *the covariance of an efficient estimator with its difference from an inefficient estimator is zero*, which implies that

$$\text{Cov}[(\mathbf{b} - \hat{\beta}), \hat{\beta}] = \text{Cov}[\mathbf{b}, \hat{\beta}] - \text{Var}[\hat{\beta}] = 0$$

or that

$$\text{Cov}[\mathbf{b}, \hat{\beta}] = \text{Var}[\hat{\beta}].$$

²¹See Balestra and Nerlove (1966), Fomby, Hill, and Johnson (1984), Judge et al. (1985), Hsiao (1986), Anderson and Hsiao (1982), Nerlove (1971a), and Baltagi (1985).

²²See Macurdy (1982) and Beggs (1986).

²³See Hausman and Taylor (1981) and Chamberlain (1975).

²⁴Related results are given by Baltagi (1986).

²⁵Referring to the GLS matrix weighted average given earlier, we see that the efficient weight uses θ , whereas OLS sets $\theta = 1$.

Inserting this result in (14-35) produces the required covariance matrix for the test,

$$\text{Var}[\mathbf{b} - \hat{\beta}] = \text{Var}[\mathbf{b}] - \text{Var}[\hat{\beta}] = \Sigma. \quad (14-36)$$

The chi-squared test is based on the Wald criterion:

$$W = \chi^2[K] = [\mathbf{b} - \hat{\beta}]' \hat{\Sigma}^{-1} [\mathbf{b} - \hat{\beta}]. \quad (14-37)$$

For $\hat{\Sigma}$, we use the estimated covariance matrices of the slope estimator in the LSDV model and the estimated covariance matrix in the random effects model, excluding the constant term. Under the null hypothesis, W is asymptotically distributed as chi-squared with K degrees of freedom.

EXAMPLE 14.5 Hausman Test

The Hausman test for the fixed and random effects regressions is based on the parts of the coefficient vectors and the asymptotic covariance matrices that correspond to the slopes in the models, that is, ignoring the constant term(s). The coefficient estimates are given in Table 14.2. The two estimated asymptotic covariance matrices are

$$\text{Est. Var}[\mathbf{b}_{FE}] = \begin{bmatrix} 0.0008934 & -0.0003178 & -0.001884 \\ -0.0003178 & 0.0002310 & -0.0007685 \\ -0.001884 & -0.0007685 & 0.04067 \end{bmatrix}$$

and

$$\text{Est. Var}[\mathbf{b}_{RE}] = \begin{bmatrix} 0.0006547 & -0.0002270 & -0.001542 \\ -0.0002270 & 0.0001961 & -0.0008968 \\ -0.001542 & -0.0008968 & 0.03991 \end{bmatrix}$$

The test statistic is 3.25. The critical value from the chi-squared table with three degrees of freedom is 7.814, which is far larger than the test value. The hypothesis that the individual effects are uncorrelated with the other regressors in the model cannot be rejected. Based on the LM test, which is decisive that there are individual effects, and the Hausman test, which suggests that these effects are uncorrelated with the other variables in the model, we would conclude that of the two alternatives we have considered, the random effects model is the better choice.

14.4.5. UNBALANCED PANELS AND RANDOM EFFECTS

Unbalanced panels add a new layer of difficulty in the random effects model. The first problem can be seen in (14-21). The matrix \mathbf{V} is no longer $\mathbf{I} \otimes \Omega$ because the diagonal blocks in \mathbf{V} are of different sizes. There is also groupwise heteroscedasticity, because the i th diagonal block in $\mathbf{V}^{-1/2}$ is

$$\Omega_i^{-1/2} = \mathbf{I}_{T_i} - \frac{\theta_i}{T_i} \mathbf{1}\mathbf{1}',$$

$$\theta_i = 1 - \frac{\sigma_\epsilon}{\sqrt{T_i \sigma_u^2 + \sigma_\epsilon^2}}.$$

In principle, estimation is still straightforward, since the source of the groupwise heteroscedasticity is only the unequal group sizes. Thus, for GLS, or FGLS with estimated

variance components, it is necessary only to use the group specific θ_i in the transformation in (14-22).

The problem arises in the estimation of the variance components. The LSDV estimator (properly computed) still provides a consistent estimator of σ_ϵ^2 . But we need a second equation to estimate σ_u^2 . We used the group means estimator for this step in (14-30). In the group means regression (14-30), the disturbances are now heteroscedastic:

$$\text{Var} \left[u_i + \frac{\sum_{t=1}^{T_i} \epsilon_{it}}{T_i} \right] = \sigma_u^2 + \frac{\sigma_\epsilon^2}{T_i} = \kappa_i^2.$$

The unbiasedness result for estimating the variance in this regression no longer holds. The OLS slope estimator for the full sample is still unbiased, however. And, we have seen a similar case in Section 12.2.2. The disturbance variance estimator in a heteroscedastic regression is a consistent estimator of

$$\bar{\kappa}^2 = \text{plim} \frac{1}{n} \sum_{i=1}^n \kappa_i^2,$$

assuming that the probability limit exists. In fact, the mean squared residual using group means is a consistent estimator of $\bar{\kappa}^2$ based on *any* consistent estimator of β . Now, in this setting, consistency still applies to n increasing, not T_i . Therefore, the variance estimator in the group means regression is a consistent estimator of

$$\begin{aligned} \bar{\kappa}^2 &= \sigma_u^2 + \sigma_\epsilon^2 \text{plim} \frac{1}{n} \sum_{i=1}^n \frac{1}{T_i} \\ &= \sigma_u^2 + \sigma_\epsilon^2 \text{plim} Q_n, \end{aligned}$$

assuming that Q_n has a probability limit (or an ordinary limit). It appears that some assumption about the group sizes is necessary. If T_i were randomly distributed across individuals around a mean of T , then $\text{plim} Q_n = 1/T$ (of course). This assumption may be realistic, but if not, then some characterization of the sequence $\{T_i\}$ is needed to claim consistency of the variance components. If we assume only that the probability limit exists and that we are estimating it consistently, then the feasible counterpart to (14-31) and (14-32) would be

$$\begin{aligned} \hat{\sigma}_{**}^2 &= \frac{\mathbf{e}'_{**} \mathbf{e}_{**}}{n - K}, \\ \hat{\sigma}_u^2 &= \hat{\sigma}_{**}^2 - \hat{\sigma}_\epsilon^2 Q_n. \end{aligned}$$

We can now continue with FGLS estimation.

14.5. Heteroscedasticity and Robust Covariance Estimation

Since the models considered here are extensions of the classical regression model, we can treat heteroscedasticity in the same way that we did in Chapter 12. That is, we can compute the ordinary or feasible generalized least squares estimators and

obtain an appropriate robust covariance matrix estimator, or we can impose some structure on the disturbance variances and use generalized least squares. In the panel data settings, there is greater flexibility for the second of these without making strong assumptions about the nature of the heteroscedasticity. We will discuss this model under the heading of "covariance structures" in Section 15.2. In this section, we will consider robust estimation of the asymptotic covariance matrix for least squares.

14.5.1. ROBUST ESTIMATION OF THE FIXED EFFECTS MODEL

In the fixed effects model, the full regressor matrix is $\mathbf{Z} = [\mathbf{D}, \mathbf{X}]$. The White heteroscedasticity consistent covariance matrix for OLS—that is, for the fixed effects estimator—is the lower right block of the partitioned matrix

$$\text{Est. Var}[\mathbf{a}, \mathbf{b}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}^2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1},$$

where \mathbf{E} is a diagonal matrix of least squares (fixed effects estimator) residuals. This computation promises to be formidable, but fortunately, it works out very simply. The White estimator for the slopes is obtained just by using the data in group mean deviation form [see (14-4) and (14-8)] in the familiar computation of \mathbf{S}_0 [see (12-7) to (12-9)]. Also, the disturbance variance estimator in (14-8) is the counterpart to the one in (12-3), which we showed that after the appropriate scaling of Ω was a consistent estimator of $\sigma^2 = \text{plim}[1/(nT)] \sum_{i=1}^n \sum_{t=1}^T \sigma_{it}^2$. The implication is that we may still use (14-8) to estimate the variances of the fixed effects.

A somewhat less general but useful simplification of this result can be obtained if we assume that the disturbance variance is constant within the i th group. If $E[\epsilon_{it}^2] = \sigma_i^2$, then, with a panel of data, σ_i^2 is estimable by $\mathbf{e}_i'\mathbf{e}_i/T$ using the least squares residuals. (This heteroscedastic regression model was considered at various points in Chapter 12.) The center matrix in $\text{Est. Var}[\mathbf{a}, \mathbf{b}]$ may be replaced with $\sum_i (\mathbf{e}_i'\mathbf{e}_i/T)\mathbf{Z}_i'\mathbf{Z}_i$. Whether this estimator is preferable is unclear. If the groupwise model is correct, then it and the White estimator will converge to the same matrix. On the other hand, if the disturbance variances do vary within the groups, then this revised computation may be inappropriate.

Arellano (1987) has taken this analysis a step further. If one takes the i th group as a whole, then we can treat the observations in

$$\mathbf{y}_i = \alpha_i \mathbf{i} + \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i$$

as a generalized regression model with disturbance covariance matrix Ω_i . We saw in Section 11.4 that a model this general, with no structure on Ω , offered little hope for estimation, robust or otherwise. But the problem is more manageable with a panel data set. As before, let \mathbf{X}_{i*} denote the data in group mean deviation form. The counterpart to $\mathbf{X}'\Omega\mathbf{X}$ here is

$$\mathbf{X}_{*}'\Omega\mathbf{X}_{*} = \sum_{i=1}^n (\mathbf{X}_{i*}'\Omega_i\mathbf{X}_{i*}).$$

By the same reasoning that we used to construct the White estimator in Chapter 12, we can consider estimating Ω_i with the sample of one, $\mathbf{e}_i\mathbf{e}_i'$. As before, it is not

consistent estimation of the individual Ω_i 's that is at issue, but estimation of the sum. If n is large enough, then we could argue that

$$\begin{aligned}\text{plim} \frac{1}{nT} \mathbf{X}'_* \Omega \mathbf{X}_* &= \text{plim} \frac{1}{nT} \sum_{i=1}^n \mathbf{X}'_{i*} \Omega_i \mathbf{X}_{i*} \\ &= \text{plim} \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \mathbf{X}'_{i*} \mathbf{e}_i \mathbf{e}_i' \mathbf{X}_{i*} \\ &= \text{plim} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e_{it} e_{is} \mathbf{x}_{it} \mathbf{x}_{is}' \right).\end{aligned}$$

The result is a combination of the White and Newey–West estimators. But the weights in the latter are 1 rather than $[1 - t/(L+1)]$ because there is no correlation across the groups, so the sum is actually just an average of finite matrices.

14.5.2. HETEROSCEDASTICITY IN THE RANDOM EFFECTS MODEL

Since the random effects model is a generalized regression model with a known structure, robust estimation of the covariance matrix for the OLS estimator in this context is not the best use of the data. If a perfectly general covariance structure is assumed, then one can simply use the results in the preceding section with a single overall constant term rather than a set of fixed effects. But within the setting of the random effects model, $w_{it} = \epsilon_{it} + u_i$, allowing the disturbance variance of the group-specific component u_i to vary across groups would be a useful extension. For estimation, we can use the following strategy. In Section 14.4.4, we introduced heteroscedasticity into estimation of the random effects model by allowing the group sizes to vary. But the estimator there (and its feasible counterpart in the next section) would be the same if, instead of $\theta_i = 1 - \sigma_\epsilon^2 / (T_i \sigma_u^2 + \sigma_\epsilon^2)^{1/2}$, we were faced with

$$\theta_i = 1 - \frac{\sigma_\epsilon}{\sqrt{T \sigma_{ui}^2 + \sigma_\epsilon^2}}$$

or even if both T and σ_u^2 varied across groups. Therefore, for computing the appropriate feasible generalized least squares estimator, once again we need only devise an estimator for the variance components and then apply the GLS transformation shown earlier. If

$$\text{Var}[w_{it}] = \sigma_\epsilon^2 + \sigma_{ui}^2,$$

then in (14-29), s^2 , the residual variance in the LSDV model, still provides a consistent estimator of σ_ϵ^2 . Within each group, we can estimate $\sigma_\epsilon^2 + \sigma_{ui}^2$ with the residual variance based on any consistent estimator of $[\alpha, \beta]$. The ordinary least squares estimator is a natural candidate, so

$$\overline{\sigma_{ui}^2 + \sigma_\epsilon^2} = \frac{\sum_{i=1}^T (e_{it} - \bar{e}_i)^2}{T - 1} = s_i^2.$$

Thereafter, an estimator of σ_{ui}^2 is

$$\hat{\sigma}_{ui}^2 = s_i^2 - s^2.$$

We can now compute the FGLS estimator as before.

There is a complication in this method that is likely to be quite common. Nothing in this prescription prevents the variance estimator from being negative. Since T (or T_i) is likely to be quite small although the full sample is likely to be large, there will be a large amount of sampling variability in s_i^2 that is averaged out (over n) of s^2 , so the difference is likely to be negative in some applications. Various patches have been suggested for this case. An expedient suggested by Baltagi (see his page 79, Table 5.1) is simply to replace negative values with zeros. This method will imply that $\hat{\theta}_i = 0$, so in the computation of the FGLS estimates, the data for this group will enter the sum of squares or cross products untransformed.

EXAMPLE 14.6 Heteroscedasticity Consistent Estimation

The fixed effects estimates for the cost equation are shown in Table 14.2. The row of standard errors labeled White (1) are the estimates based on the usual calculation. For two of the three coefficients, these are actually substantially smaller than the least squares results. The estimates labeled White (2) are based on the groupwise heteroscedasticity model suggested earlier. These estimates are essentially the same as White (1). As noted, it is unclear whether this computation is preferable. Of course, if it were known that the groupwise model were correct, then the least squares computation itself would be inefficient and, in any event, a two-step FGLS estimator would be better.

The estimator of σ_u^2 based on the least squares residuals is 0.0036125. The six individual estimators of σ_u^2 are 0.04023², 0.06889², 0.05161², 0.09524², 0.04817², and 0.06306². Three of the six estimators for σ_{ui}^2 are negative based on these results, which suggests that a groupwise heteroscedastic random effects model is not an appropriate specification for these data.

14.6. Autocorrelation

Autocorrelation in the fixed effects model is a minor extension of the model of the preceding chapter. With the LSDV estimator in hand, estimates of the parameters of a disturbance process and transformations of the data to allow FGLS estimation proceed exactly as before. The extension one might consider is to allow the autocorrelation coefficient(s) to vary across groups. But even if so, treating each group of observations as a sample in itself provides the appropriate framework for estimation.

In the random effects model, as before, there are additional complications. The regression model is

$$y_{it} = \alpha + \beta'x_{it} + \epsilon_{it} + u_i.$$

If ϵ_{it} is produced by an AR(1) process, $\epsilon_{it} = \rho\epsilon_{i,t-1} + v_{it}$, then the familiar partial differencing procedure we used before would produce²⁶

$$\begin{aligned} y_{it} - \rho y_{i,t-1} &= \alpha(1 - \rho) + \beta'(x_{it} - \rho x_{i,t-1}) + \epsilon_{it} - \rho\epsilon_{i,t-1} + u_i(1 - \rho) \\ &= \alpha(1 - \rho) + \beta'(x_{it} - \rho x_{i,t-1}) + v_{it} + u_i(1 - \rho) \\ &= \alpha(1 - \rho) + \beta'(x_{it} - \rho x_{i,t-1}) + v_{it} + w_i. \end{aligned}$$

Therefore, if an estimator of ρ were in hand, then one could at least treat partially differenced observations 2 – T in each group as the same random effects model that we just examined. Variance estimators would have to be adjusted by a factor of $(1 - \rho)^2$. Two issues remain: (1) how is the estimate of ρ obtained and (2) how does one treat the first observation? For the first of these, the first autocorrelation coefficient of the LSDV residuals (so as to purge the residuals of the individual specific effects, u_i) is a simple expedient. This estimator will be consistent in nT . It is in T alone, but, of course, T is likely to be small. The second question is more difficult. Estimation is simple if the first observation is simply dropped. We saw in Chapter 13 that omitting the first observation in a time series could lead to a serious loss of efficiency. If the number of cross-section units is small, then the same effect might arise here. But if the panel contains many groups (large n), then omitting the first observation is less likely to cause the same kinds of problems. One can apply the Prais–Winsten transformation to the first observation in each group instead [multiply by $(1 - \rho^2)^{1/2}$], but then an additional complication arises at the second (FGLS) step when the observations are transformed a second time. On balance, the Cochrane–Orcutt estimator is probably a reasonable middle ground. Baltagi (1995, p. 83) discusses the procedure. He also discusses estimation in higher-order AR and MA processes.

In the same manner as in the previous section, we could allow the autocorrelation to differ across groups. An estimate of each ρ_i is computable using the group mean deviation data. This estimator is consistent in T , which is problematic in this setting. In the earlier case, we overcame this difficulty by averaging over n such “weak” estimates and achieving consistency in the dimension of n instead. We lose that advantage when we allow ρ to vary over the groups. This result is the same that arose in our treatment of heteroscedasticity.

For the airlines data in our examples, the estimated autocorrelation is 0.5086, which is fairly large. Estimates of the fixed and random effects models using the Cochrane–Orcutt procedure for correcting the autocorrelation are given in Table 14.2. Despite the large value of r , the resulting changes in the parameter estimates and standard errors are quite modest.

14.7. Dynamic Models

Panel data are well suited for examining dynamic effects, as in the first-order model,

$$y_{it} = \alpha_i + x'_{it}\beta + \delta y_{i,t-1} + \epsilon_{it}.$$

²⁶See Lillard and Willis (1978).

Substantial complications arise in estimation of such a model, however. In both the fixed and random effects settings, the difficulty is that the lagged dependent variable is correlated with the disturbance, even if it is assumed that ϵ_{it} is not itself autocorrelated.

For the moment, we can think of the fixed effects model as an ordinary regression with a lagged variable. We considered this case in Section 9.4.3 as a regression with a stochastic regressor that is dependent across observations. In the dynamic regression model, the estimator based on T observations is not unbiased, but it is consistent in T . That conclusion was the main result of Section 9.4.3. The finite sample bias is of order $1/T$. The same result applies here, but the difference is that whereas before we obtained our large sample results by allowing T to grow large, in this setting, T is assumed to be small, and large-sample results are obtained with respect to n growing large, not T . The fixed effects estimator of $\theta = [\beta, \delta]$ can be viewed as an average of n estimators. For example, if $T \geq K$, then, from (14-4),

$$\begin{aligned}\hat{\theta} &= \left[\sum_{i=1}^n \mathbf{X}_i' \mathbf{M}_d \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i' \mathbf{M}_d \mathbf{y}_i \right] \\ &= \left[\sum_{i=1}^n \mathbf{X}_i' \mathbf{M}_d \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{X}_i' \mathbf{M}_d \mathbf{X}_i \mathbf{b}_i \right] \\ &= \sum_{i=1}^n \mathbf{W}_i \mathbf{b}_i.\end{aligned}$$

The average of n inconsistent estimators will still be inconsistent. (This analysis is only heuristic. If $T < K$, then the individual coefficient vectors cannot be computed.²⁷) The problem is more transparent in the random effects model. The lagged variable is correlated with the compound disturbance in the model, since u_i (which is α_i) enters the equation for every observation in group i . Neither of these results renders the model inestimable, but they do make some technique other than LSDV or FGLS necessary.

The general approach, which has been developed in several stages in the literature,²⁸ relies on instrumental variables estimators and, most recently [by Ahn and Schmidt (1993)] on a GMM estimator. In either the fixed or random effects cases, the heterogeneity can be swept from the model by taking first differences, which produces

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \beta + \delta(y_{i,t-1} - y_{i,t-2}) + (\epsilon_{it} - \epsilon_{i,t-1}).$$

This model is still complicated by correlation between the lagged dependent variable and the disturbance (and by its first-order moving average disturbance). But without the group effects, there is a simple instrumental variables estimator available. Assuming that the time series is long enough, one could use the differences, $(y_{i,t-2} - y_{i,t-3})$, or the lagged levels, $y_{i,t-2}$ and $y_{i,t-3}$, as one or two instrumental variables for

²⁷Further discussion is given by Nickell (1981), Ridder and Wansbeek (1990), and Kiviet (1995).

²⁸See, for example, Anderson and Hsiao (1981), Arellano (1989), Arellano and Bond (1991), Arellano and Bover (1993), and Ahn and Schmidt (1993).

$(y_{it-1} - y_{it-2})$. (The other variables can serve as their own instruments.) By this construction, then, the treatment of this model is a standard application of the instrumental variables technique that we developed in Section 9.5. There is a question as to whether one should use differences or levels as instruments. Arellano (1989) gives evidence that the latter is preferable.

Ahn and Schmidt (among others) observed that the IV estimator neglects quite a lot of information and is therefore inefficient. For example, in the first differenced model,

$$E[y_{it}, (\epsilon_{it} - \epsilon_{it-1})] = 0, \quad s = 0, \dots, t-2, t = 2, \dots, T.$$

That is, the *level* of y is uncorrelated with the differences of disturbances that are at least two periods subsequent. The corresponding moment equations that can enter the construction of a GMM estimator are

$$\frac{1}{n} \sum_{i=1}^n y_{is} [(y_{it} - y_{it-1}) - (\mathbf{x}_{it} - \mathbf{x}_{it-1})' \boldsymbol{\beta} - \delta(y_{it-1} - y_{it-2})] = 0$$

$$s = 0, \dots, t-2, \quad t = 2, \dots, T.$$

Altogether, Ahn and Schmidt identify $T(T-1)/2 + T - 2$ such equations that involve mixtures of the levels and differences of the variables. The main conclusion that they demonstrate is that in the dynamic model, there is a large amount of information to be culled not only from the familiar relationships among the levels of the variables but also from the implied relationships between the levels and the first differences.

14.8. Conclusions

The preceding has shown a few of the extensions of the classical model that can be obtained when panel data are available. In principle, any of the models we have examined before this chapter and all those we will consider later, including the multiple equation models, can be extended in the same way. The main advantage, as we noted at the outset, is that with panel data, one can formally model the heterogeneity across groups that is typical in microeconomic data.

We will find in Chapter 15 that to some extent this model of heterogeneity can be misleading. What might have appeared at one level to be differences in the variances of the disturbances across groups may well be due to heterogeneity of a different sort, associated with the coefficient vectors. We will consider this possibility in the next chapter. We will also examine some additional models for disturbance processes that arise naturally in a multiple equations context but are actually more general cases of some of the models we looked at above, such as the model of groupwise heteroscedasticity.

Exercises

1. The following is a panel of data on investment (y) and profit (x) for $n = 3$ firms over $T = 10$ periods.

random vector is defined to be the matrix of variances and covariances of its elements. We write

$$E(\mathbf{y}) = \boldsymbol{\mu}, \quad V(\mathbf{y}) = \boldsymbol{\Sigma}.$$

When \mathbf{y} is $n \times 1$, then $\boldsymbol{\mu}$ is $n \times 1$, and $\boldsymbol{\Sigma}$ is $n \times n$ symmetric.

Let $\boldsymbol{\epsilon} = \mathbf{y} - \boldsymbol{\mu} = \{y_i - \mu_i\}$ be the $n \times 1$ vector of deviations of the y 's from their respective expectations. So

$$\boldsymbol{\epsilon}\boldsymbol{\epsilon}' = (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})' = \{(y_h - \mu_h)(y_i - \mu_i)\}$$

is an $n \times n$ symmetric random matrix whose elements are the squares and cross-products of those deviations. Then

$$E(\boldsymbol{\epsilon}) = E(\mathbf{y} - \boldsymbol{\mu}) = \{E(y_i) - \mu_i\} = \{\mu_i - \mu_i\} = \{0\} = \mathbf{0},$$

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'] = \{\sigma_{hi}\} = \boldsymbol{\Sigma} = V(\mathbf{y}) = V(\boldsymbol{\epsilon}).$$

The *covariance matrix of a pair of random vectors* is defined to be the matrix of covariances between the elements of one vector and the elements of the other vector. Thus if $\mathbf{z} = \{z_h\}$ is an $m \times 1$ random vector and $\mathbf{y} = \{y_i\}$ is an $n \times 1$ random vector, then

$$C(\mathbf{z}, \mathbf{y}) = E\{[\mathbf{z} - E(\mathbf{z})][\mathbf{y} - E(\mathbf{y})]'\}$$

is the $m \times n$ matrix whose (h, i) th element is $C(z_h, y_i)$, while $C(\mathbf{y}, \mathbf{z})$ is the $n \times m$ transpose of that matrix.

Here are a set of rules for calculating expectations, variances, and covariances of certain functions of \mathbf{y} . Throughout we suppose that the $n \times 1$ random vector \mathbf{y} has expectation vector $E(\mathbf{y}) = \boldsymbol{\mu}$ and variance matrix $V(\mathbf{y}) = \boldsymbol{\Sigma}$, and write $\boldsymbol{\epsilon} = \mathbf{y} - \boldsymbol{\mu}$. The first two rules, which refer to linear functions, are straightforward generalizations of T5 and T6 in Section 5.1.

R1. SCALAR LINEAR FUNCTION. Let $z = g + \mathbf{h}'\mathbf{y}$, where the scalar g and the $n \times 1$ vector \mathbf{h} are constants. Then the random variable z has

$$E(z) = g + \mathbf{h}'E(\mathbf{y}) = g + \mathbf{h}'\boldsymbol{\mu}.$$

Further, let $z^* = z - E(z)$. Then $z^* = \mathbf{h}'\mathbf{y} - \mathbf{h}'\boldsymbol{\mu} = \mathbf{h}'(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{h}'\boldsymbol{\epsilon}$ and $z^{*2} = (\mathbf{h}'\boldsymbol{\epsilon})^2 = (\mathbf{h}'\boldsymbol{\epsilon})(\mathbf{h}'\boldsymbol{\epsilon}) = (\mathbf{h}'\boldsymbol{\epsilon})(\boldsymbol{\epsilon}'\mathbf{h}) = \mathbf{h}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{h}$. So

$$V(z) = E(z^{*2}) = E(\mathbf{h}'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{h}) = \mathbf{h}'E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{h} = \mathbf{h}'V(\boldsymbol{\epsilon})\mathbf{h} = \mathbf{h}'\boldsymbol{\Sigma}\mathbf{h}.$$

15.1. Matrix Algebra for Random Variables

In this chapter we will establish a model for multiple regression, that is, a population specification and sampling scheme that support running, LS linear regression to estimate population parameters. In preparation, we develop a general matrix-algebra system for dealing with random variables.

Setting aside the regression application, let Y_1, \dots, Y_n be a set of n random variables whose joint pdf (or pmf) is $f(y_1, \dots, y_n)$. The expectations, variances, and covariances are (for $i, h = 1, \dots, n$):

$$E(Y_i) = \mu_i, \quad V(Y_i) = \sigma_i^2 = \sigma_{ii}, \quad C(Y_h, Y_i) = \sigma_{hi} = \sigma_{ih}.$$

It is natural to display these in an $n \times 1$ vector \mathbf{Y} , an $n \times 1$ vector $\boldsymbol{\mu}$, and an $n \times n$ matrix $\boldsymbol{\Sigma}$, where:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}.$$

At the risk of some confusion we adopt the matrix-algebra convention of lowercase characters for vectors, overriding the statistical convention of uppercase characters for random variables, and write the $n \times 1$ random vector and its elements as

$$\mathbf{y} = (y_1, \dots, y_n)'.$$

The *expectation of a random vector* (or matrix) is defined to be the vector (or matrix) of expectations of its elements. The *variance matrix* of a

Incidentally, since the scalar variance $V(z)$ must be nonnegative, we see that every variance matrix Σ is nonnegative definite, and is positive definite iff it is nonsingular.

R2. VECTOR LINEAR FUNCTION. Let $\mathbf{z} = \mathbf{g} + \mathbf{H}\mathbf{y}$, where the $k \times 1$ vector \mathbf{g} and the $k \times n$ matrix \mathbf{H} are constants. Then the $k \times 1$ random vector \mathbf{z} has

$$E(\mathbf{z}) = \mathbf{g} + \mathbf{H}\boldsymbol{\mu}.$$

Further, let $\mathbf{z}^* = \mathbf{z} - E(\mathbf{z})$. Then $\mathbf{z}^* = \mathbf{H}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{H}\boldsymbol{\epsilon}$, and $\mathbf{z}^*\mathbf{z}^{*'} = \mathbf{H}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{H}'$. So

$$V(\mathbf{z}) = E(\mathbf{z}^*\mathbf{z}^{*'}) = E(\mathbf{H}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{H}') = \mathbf{H}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{H}' = \mathbf{H}\Sigma\mathbf{H}'.$$

R3. MEAN SQUARES. Let $\mathbf{W} = \mathbf{y}\mathbf{y}'$. Then the $n \times n$ random matrix \mathbf{W} has expectation $E(\mathbf{W}) = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}'$.

Proof. Write

$$\mathbf{y}\mathbf{y}' = (\boldsymbol{\mu} + \boldsymbol{\epsilon})(\boldsymbol{\mu} + \boldsymbol{\epsilon})' = \boldsymbol{\mu}\boldsymbol{\mu}' + \boldsymbol{\mu}\boldsymbol{\epsilon}' + \boldsymbol{\epsilon}\boldsymbol{\mu}' + \boldsymbol{\epsilon}\boldsymbol{\epsilon}',$$

which, since $\boldsymbol{\mu}$ is constant and $E(\boldsymbol{\epsilon}) = \mathbf{0}$, implies

$$E(\mathbf{y}\mathbf{y}') = \boldsymbol{\mu}\boldsymbol{\mu}' + \Sigma. \quad \blacksquare$$

R4. SUM OF SQUARES. Let $w = \mathbf{y}'\mathbf{y}$. Then the scalar random variable w has expectation $E(w) = \text{tr}(\Sigma) + \boldsymbol{\mu}'\boldsymbol{\mu}$.

Proof. Write

$$\mathbf{y}'\mathbf{y} = \text{tr}(\mathbf{y}'\mathbf{y}) = \text{tr}(\mathbf{y}\mathbf{y}') = \text{tr}(\mathbf{W}),$$

so

$$\begin{aligned} E(\mathbf{y}'\mathbf{y}) &= E[\text{tr}(\mathbf{W})] = \text{tr}[E(\mathbf{W})] = \text{tr}(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}') \\ &= \text{tr}(\Sigma) + \text{tr}(\boldsymbol{\mu}\boldsymbol{\mu}') = \text{tr}(\Sigma) + \text{tr}(\boldsymbol{\mu}'\boldsymbol{\mu}) = \text{tr}(\Sigma) + \boldsymbol{\mu}'\boldsymbol{\mu}, \end{aligned}$$

using the facts that trace is a linear operator, and that if \mathbf{AB} and \mathbf{BA} are both square matrices, then $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. \blacksquare

R5. QUADRATIC FORM. Let $w = \mathbf{y}'\mathbf{T}\mathbf{y}$, where the $n \times n$ matrix \mathbf{T} is constant. Then the random variable w has expectation $E(w) = \text{tr}(\mathbf{T}\Sigma) + \boldsymbol{\mu}'\mathbf{T}\boldsymbol{\mu}$.

Proof. Write $\mathbf{y}'\mathbf{T}\mathbf{y} = \text{tr}(\mathbf{y}'\mathbf{T}\mathbf{y}) = \text{tr}(\mathbf{T}\mathbf{y}\mathbf{y}') = \text{tr}(\mathbf{T}\mathbf{W})$. Then

$$\begin{aligned} E(\mathbf{y}'\mathbf{T}\mathbf{y}) &= E[\text{tr}(\mathbf{T}\mathbf{W})] = \text{tr}[E(\mathbf{T}\mathbf{W})] = \text{tr}[\mathbf{T}E(\mathbf{W})] \\ &= \text{tr}[\mathbf{T}(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}')] = \text{tr}(\mathbf{T}\Sigma) + \text{tr}(\mathbf{T}\boldsymbol{\mu}\boldsymbol{\mu}') \\ &= \text{tr}(\mathbf{T}\Sigma) + \boldsymbol{\mu}'\mathbf{T}\boldsymbol{\mu}. \quad \blacksquare \end{aligned}$$

R6. PAIR OF VECTOR LINEAR FUNCTIONS. Let $\mathbf{z}_1 = \mathbf{g}_1 + \mathbf{H}_1\mathbf{y}$, $\mathbf{z}_2 = \mathbf{g}_2 + \mathbf{H}_2\mathbf{y}$, where the $m_1 \times 1$ vector \mathbf{g}_1 , the $m_2 \times 1$ vector \mathbf{g}_2 , the $m_1 \times n$ matrix \mathbf{H}_1 , and the $m_2 \times n$ matrix \mathbf{H}_2 are constants. Then $C(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{H}_1\Sigma\mathbf{H}_2'$.

Proof. Let $\mathbf{z}_1^* = \mathbf{z}_1 - E(\mathbf{z}_1) = \mathbf{H}_1\boldsymbol{\epsilon}$, and $\mathbf{z}_2^* = \mathbf{z}_2 - E(\mathbf{z}_2) = \mathbf{H}_2\boldsymbol{\epsilon}$. Then $\mathbf{z}_1^*\mathbf{z}_2^{*'} = \mathbf{H}_1\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{H}_2'$, so

$$C(\mathbf{z}_1, \mathbf{z}_2) = E(\mathbf{z}_1^*\mathbf{z}_2^{*'}) = \mathbf{H}_1E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{H}_2' = \mathbf{H}_1\Sigma\mathbf{H}_2'. \quad \blacksquare$$

15.2. Classical Regression Model

We now set out the statistical model that is most commonly used to justify running a sample LS regression to estimate population parameters. That is, we provide a context for the data, one in which we observe a drawing on an $n \times 1$ random vector \mathbf{y} and an $n \times k$ matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$. The *classical regression*, or CR, model consists of these four assumptions:

$$(15.1) \quad E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta},$$

$$(15.2) \quad V(\mathbf{y}) = \sigma^2\mathbf{I},$$

$$(15.3) \quad \mathbf{X} \text{ nonstochastic},$$

$$(15.4) \quad \text{rank}(\mathbf{X}) = k.$$

The understanding is that we observe \mathbf{X} and \mathbf{y} , while $\boldsymbol{\beta}$ and σ^2 are unknown.

We interpret the assumptions briefly. In general, an $n \times 1$ random vector $\mathbf{y} = (y_1, \dots, y_n)'$ will have expectation vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_n \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \vdots & \sigma_{ni} & \vdots \\ \vdots & \vdots & \ddots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}.$$

So in general, the elements of a random vector \mathbf{y} will have different expectations, different variances, and free covariances. But in the CR model, we have

$$(15.1) \quad \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta},$$

which says that $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$, where \mathbf{x}_i' is the i th row of \mathbf{X} . (Caution: Do not confuse \mathbf{x}_i' with the transpose of the i th column of \mathbf{X} .) Consequently, all n of the unknown expectations, the μ_i 's, are expressible in terms of k unknown parameters, the β_j 's. The n expectations may well be different, but they all lie in the same k -dimensional plane in n -space. Further, in the CR model, we have

$$(15.2) \quad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I},$$

which says that $\sigma_{ii} = \sigma^2$ for all i , and that $\sigma_{ii} = 0$ for all $i \neq j$. Thus the random variables y_1, \dots, y_n all have the same variance, and are uncorrelated. Further, we have

$$(15.3) \quad \mathbf{X} \text{ nonstochastic,}$$

which says that the elements of \mathbf{X} are constants, that is, degenerate random variables. Their values are fixed in repeated samples, unlike the elements of \mathbf{y} which, being random variables, will vary from sample to sample. Finally, we have

$$(15.4) \quad \text{rank}(\mathbf{X}) = k,$$

which says that the $n \times k$ matrix \mathbf{X} has full column rank; its k columns are linearly independent in the matrix algebra sense.

In Chapter 16, we will return to the interpretation of the CR model, and to the population and sampling assumptions that underlie it.

15.3. Estimation of $\boldsymbol{\beta}$

We proceed to the estimation of the unknown parameters $\boldsymbol{\beta}$ and σ^2 . We have a sample (\mathbf{y}, \mathbf{X}) produced by the CR model. How shall we process the sample data to obtain parameter estimates? The proposal is to use the sample LP, that is, to run the LS linear regression of \mathbf{y} on \mathbf{X} . Because $\text{rank}(\mathbf{X}) = k$, the normal equations of LS linear regression will have a unique solution, namely

$$\mathbf{b} = \mathbf{A}\mathbf{y}, \quad \text{where } \mathbf{A} = \mathbf{Q}^{-1}\mathbf{X}'.$$

This $k \times 1$ random vector \mathbf{b} is our estimator of $\boldsymbol{\beta}$.

What properties does the estimator have? The matrix \mathbf{A} is constant because it is a function of \mathbf{X} alone. Hence \mathbf{b} is a linear function of the random vector \mathbf{y} , and R2 of Section 15.1 applies. Recalling that $\mathbf{A}\mathbf{X} = \mathbf{I}$ and that $\mathbf{A}\mathbf{A}' = \mathbf{Q}^{-1}$, we calculate

$$E(\mathbf{b}) = \mathbf{A}E(\mathbf{y}) = \mathbf{A}\boldsymbol{\mu} = \mathbf{A}(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{A}\mathbf{X})\boldsymbol{\beta} = \mathbf{I}\boldsymbol{\beta} = \boldsymbol{\beta},$$

$$V(\mathbf{b}) = \mathbf{A}V(\mathbf{y})\mathbf{A}' = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \mathbf{A}(\sigma^2\mathbf{I})\mathbf{A}' = \sigma^2\mathbf{A}\mathbf{A}' = \sigma^2\mathbf{Q}^{-1}.$$

So the LS coefficient vector \mathbf{b} is an unbiased estimator of the parameter vector $\boldsymbol{\beta}$, with $E(b_j) = \beta_j$ for $j = 1, \dots, k$. And the variances and covariances of the k random variables in \mathbf{b} are given by the appropriate elements of the $k \times k$ matrix $\sigma^2\mathbf{Q}^{-1}$:

$$V(b_j) = \sigma^2 q^{jj}, \quad C(b_h, b_j) = \sigma^2 q^{hj},$$

where q^{hj} denotes the element in the h th row and j th column of \mathbf{Q}^{-1} .

15.4. Gauss-Markov Theorem

We now show that the LS estimator has an optimality property.

GAUSS-MARKOV THEOREM. In the CR model, the LS coefficient vector \mathbf{b} is the minimum variance linear unbiased estimator of the parameter vector $\boldsymbol{\beta}$.

Proof. Let $\mathbf{b}^* = \mathbf{A}^*\mathbf{y}$, where \mathbf{A}^* is any $k \times n$ nonstochastic matrix. Then \mathbf{b}^* is a linear function of \mathbf{y} , that is a linear estimator. Rule R2 gives

$$E(\mathbf{b}^*) = \mathbf{A}^* E(\mathbf{y}) = \mathbf{A}^* \mathbf{X} \boldsymbol{\beta},$$

$$V(\mathbf{b}^*) = \mathbf{A}^* V(\mathbf{y}) \mathbf{A}^{*'} = \sigma^2 \mathbf{A}^* \mathbf{A}^{*'}.$$

Clearly \mathbf{b}^* will be unbiased for $\boldsymbol{\beta}$ iff $\mathbf{A}^* \mathbf{X} = \mathbf{I}$. Write $\mathbf{A}^* = \mathbf{A} + \mathbf{D}$, where $\mathbf{A} = \mathbf{Q}^{-1} \mathbf{X}'$ and $\mathbf{D} = \mathbf{A}^* - \mathbf{A}$. Observe that

$$\mathbf{A}^* \mathbf{X} = \mathbf{A} \mathbf{X} + \mathbf{D} \mathbf{X} = \mathbf{I} + \mathbf{D} \mathbf{X},$$

$$\mathbf{A}^* \mathbf{A}^{*'} = (\mathbf{A} + \mathbf{D})(\mathbf{A} + \mathbf{D})' = \mathbf{A} \mathbf{A}' + \mathbf{D} \mathbf{D}' + \mathbf{A} \mathbf{D}' + \mathbf{D} \mathbf{A}'.$$

So the unbiasedness condition $\mathbf{A}^* \mathbf{X} = \mathbf{I}$ is equivalent to $\mathbf{D} \mathbf{X} = \mathbf{O}$, that is, to $\mathbf{D} \mathbf{X} \mathbf{Q}^{-1} = \mathbf{O}$, that is, to $\mathbf{D} \mathbf{A}' = \mathbf{O}$ (and hence to $\mathbf{A} \mathbf{D}' = \mathbf{O}$). So if \mathbf{b}^* is a linear unbiased estimator of $\boldsymbol{\beta}$, then

$$V(\mathbf{b}^*) = \sigma^2 (\mathbf{A} \mathbf{A}' + \mathbf{D} \mathbf{D}') = V(\mathbf{b}) + \sigma^2 \mathbf{D} \mathbf{D}'.$$

The matrix $\mathbf{D} \mathbf{D}'$ is nonnegative definite and the scalar σ^2 is positive, so $\sigma^2 \mathbf{D} \mathbf{D}'$ is nonnegative definite. Consequently, $V(\mathbf{b}^*) \geq V(\mathbf{b})$, with equality iff $\mathbf{D} \mathbf{D}' = \mathbf{O}$, that is, iff $\mathbf{D} = \mathbf{O}$, that is, iff $\mathbf{A}^* = \mathbf{A}$, that is, iff $\mathbf{b}^* = \mathbf{b}$ in every sample. ■

Some explanations are in order:

- The matrix $\mathbf{D} \mathbf{D}'$ is nonnegative definite because for any $k \times 1$ vector \mathbf{h} , the quadratic form $\mathbf{h}' \mathbf{D} \mathbf{D}' \mathbf{h} = (\mathbf{D}' \mathbf{h})' (\mathbf{D}' \mathbf{h}) = \mathbf{v}' \mathbf{v} \geq 0$.
- If \mathbf{t}^* and \mathbf{t} are random vectors, we say that $V(\mathbf{t}^*) \geq V(\mathbf{t})$ iff $V(\mathbf{t}^*) - V(\mathbf{t})$ is nonnegative definite.

Observe the implications of the nonnegative definiteness of $V(\mathbf{b}^*) - V(\mathbf{b})$. Element by element, \mathbf{b} is preferable to \mathbf{b}^* , any other linear unbiased estimator of $\boldsymbol{\beta}$, because its elements have smaller variances. But also consider a linear combination of the β_j 's, say $\theta = \mathbf{h}' \boldsymbol{\beta}$, where \mathbf{h} is a constant $k \times 1$ vector. Let $\mathbf{t} = \mathbf{h}' \mathbf{b}$ and let $\mathbf{t}^* = \mathbf{h}' \mathbf{b}^*$. Then both \mathbf{t} and \mathbf{t}^* are unbiased for θ , but $V(\mathbf{t}^*) - V(\mathbf{t}) = \mathbf{h}' [V(\mathbf{b}^*) - V(\mathbf{b})] \mathbf{h} \geq 0$. So \mathbf{b} is also preferable to \mathbf{b}^* for constructing estimators of linear combinations.

15.5. Estimation of σ^2 and $V(\mathbf{b})$

For estimation of the parameter σ^2 , we draw on the LS residual vector $\mathbf{e} = \mathbf{M} \mathbf{y}$. The matrix \mathbf{M} is constant because it is a function of \mathbf{X} alone. Hence $\mathbf{e} = \mathbf{M} \mathbf{y}$ is a linear function of the random vector \mathbf{y} , and \mathbf{R}^2 applies. Recalling that $\mathbf{M} \mathbf{X} = \mathbf{O}$ and $\mathbf{M} \mathbf{M}' = \mathbf{M}$, we calculate

$$E(\mathbf{e}) = \mathbf{M} E(\mathbf{y}) = \mathbf{M} (\mathbf{X} \boldsymbol{\beta}) = (\mathbf{M} \mathbf{X}) \boldsymbol{\beta} = \mathbf{O} \boldsymbol{\beta} = \mathbf{O},$$

$$V(\mathbf{e}) = \mathbf{M} V(\mathbf{y}) \mathbf{M}' = \mathbf{M} (\sigma^2 \mathbf{I}) \mathbf{M}' = \sigma^2 \mathbf{M} \mathbf{M}' = \sigma^2 \mathbf{M}.$$

Thus, considered as random variables, the residuals e_1, \dots, e_n have zero expectations, generally different variances, and nonzero covariances. Now calculate the expectation of the random variable $\mathbf{e}' \mathbf{e}$, the sum of squared residuals. Apply \mathbf{R}^4 , with \mathbf{e} playing the role of \mathbf{y} :

$$E(\mathbf{e}' \mathbf{e}) = \text{tr}[V(\mathbf{e})] + [E(\mathbf{e})]' [E(\mathbf{e})] = \text{tr}(\sigma^2 \mathbf{M}) + \mathbf{O}' \mathbf{O} = \sigma^2 \text{tr}(\mathbf{M}).$$

But $\mathbf{N} = \mathbf{X} \mathbf{A}$ and $\mathbf{A} \mathbf{X} = \mathbf{I}$, so

$$\text{tr}(\mathbf{N}) = \text{tr}(\mathbf{X} \mathbf{A}) = \text{tr}(\mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{I}_k) = k.$$

Hence for $\mathbf{M} = \mathbf{I} - \mathbf{N}$, we have

$$\text{tr}(\mathbf{M}) = \text{tr}(\mathbf{I} - \mathbf{N}) = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{N}) = n - k.$$

So

$$E(\mathbf{e}' \mathbf{e}) = \sigma^2 (n - k).$$

Defining the *adjusted mean squared residual*,

$$\hat{\sigma}^2 = \mathbf{e}' \mathbf{e} / (n - k),$$

we have $E(\hat{\sigma}^2) = E(\mathbf{e}' \mathbf{e}) / (n - k) = \sigma^2$. So $\hat{\sigma}^2$ is an unbiased estimator of σ^2 .

Finally, we estimate the variance matrix $V(\mathbf{b}) = \sigma^2 \mathbf{Q}^{-1}$, by

$$\hat{V}(\mathbf{b}) = \hat{\sigma}^2 \mathbf{Q}^{-1}.$$

Because $E(\hat{\sigma}^2) = \sigma^2$ and \mathbf{Q}^{-1} is constant, it follows that

$$E[\hat{V}(\mathbf{b})] = \sigma^2 \mathbf{Q}^{-1} = V(\mathbf{b}),$$

so that $\hat{V}(\mathbf{b})$ is an unbiased estimator of $V(\mathbf{b})$. In particular,

$$\hat{\sigma}_{b_j}^2 = \hat{\sigma}^2 q^{jj}$$

is an unbiased estimator of $V(b_j) = \sigma_{b_j}^2 = \sigma^2 q^{jj}$. The square root of the estimated variance,

$$\hat{\sigma}_{b_j} = \hat{\sigma} \sqrt{q^{jj}},$$

serves as the standard error of b_j .

Exercises

15.1 Suppose that the random vector \mathbf{x} has $E(\mathbf{x}) = \boldsymbol{\mu}$, $V(\mathbf{x}) = \boldsymbol{\Sigma}$, and that $\mathbf{y} = \mathbf{g} + \mathbf{H}\mathbf{x}$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 3 & 1 \\ 2 & 1 & 4 \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 2 & 1 \end{pmatrix}.$$

Calculate $E(\mathbf{y})$, $V(\mathbf{y})$, $E(\mathbf{y}\mathbf{y}')$, $E(\mathbf{y}'\mathbf{y})$, $C(\mathbf{y}, \mathbf{x})$, and $C(\mathbf{x}, \mathbf{y})$.

15.2 Suppose the CR model applies with $n = 40$, $\sigma^2 = 4$, and

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 40 & 10 \\ 10 & 5 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Let \mathbf{b} be the LS coefficient vector and $t = \mathbf{b}'\mathbf{b}$. Find $E(t)$.

15.3 The CR model applies with $\sigma^2 = 2$, and

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

A sample is drawn and the LS coefficients b_1 and b_2 are calculated.

- Guess, as best you can, the value of b_2 . Explain.
- Now you are told that $b_1 = 4$. Guess, as best you can, the value of b_2 . Explain.

15.4 The CR model applies along with the usual notation. For each of the following statements, indicate whether it is true or false, and justify your answer.

- The random variable $t = \mathbf{b}'\mathbf{b}$ is an unbiased estimator of the parameter $\boldsymbol{\theta} = \boldsymbol{\beta}'\boldsymbol{\beta}$.
- Since $\hat{\mathbf{y}} = \mathbf{N}\mathbf{y}$, it follows that $\mathbf{y} = \mathbf{N}^{-1}\hat{\mathbf{y}}$.
- Since $E(\hat{\mathbf{y}}) = E(\mathbf{y})$, it follows that the sum of the residuals is zero.
- If b_1 and b_2 are the first two elements of \mathbf{b} , $t_1 = b_1 + b_2$, and $t_2 = b_1 - b_2$, then $V(t_1) \geq V(t_2)$.

15.5 Show that the LS coefficients $\mathbf{b} = \mathbf{A}\mathbf{y}$ are uncorrelated with the residuals $\mathbf{e} = \mathbf{M}\mathbf{y}$. Hint: See R6, Section 15.1.

15.6 Suppose that the CR model applies to the data of Exercise 14.1. Report your estimates of the β_j parameters, with standard errors in parentheses beneath the coefficient estimates. Also report $\hat{\sigma}^2$.

of μ as the "true value" of y and of ϵ as an "error" or "mistake." For example, Judge et al. (1988, p. 179) say that the disturbance ϵ "is a random vector representing the unpredictable or uncontrollable errors associated with the outcome of the experiment," and Johnston (1984, p. 169) says that "if the theorist has done a good job in specifying all the significant explanatory variables to be included in \mathbf{X} , it is reasonable to assume that both positive and negative discrepancies from the expected value will occur and that, on balance, they will average out at zero." Such language may overdramatize the primitive concept of the difference between the observed and the expected values of a random variable. In any event, we will want to distinguish between the *disturbance* vector $\epsilon = y - \mu$, which is unobserved, and the *residual* vector $e = y - \hat{y}$, which is observed.

In what situation would the CR model be justified? Suppose that there is a multivariate population for the random vector $(y, x_2, \dots, x_k)'$, with pdf or pmf $f(y, x_2, \dots, x_k)$. Expectations, variances, and covariances are defined in the usual manner:

$$E(y) = \mu_y, \quad V(y) = \sigma_y^2, \quad C(x_h, x_j) = \sigma_{hj}, \quad C(x_j, y) = \sigma_{jy},$$

and so forth. Suppose further that the conditional expectation function of y given the x 's is linear,

$$E(y|x_2, \dots, x_k) = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

and that the conditional variance function of y given the x 's is constant,

$$V(y|x_2, \dots, x_k) = \sigma^2,$$

say. We write these compactly as

$$(16.5) \quad E(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}, \quad V(y|\mathbf{x}) = \sigma^2,$$

where $\mathbf{x} = (x_1, \dots, x_k)'$ with $x_1 = 1$, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$.

As for sampling schemes, the most natural one to consider would be:

Random Sampling from the Multivariate Population. Here n independent drawings, $(y_1, \mathbf{x}_1'), \dots, (y_n, \mathbf{x}_n')$, are made, giving the observed sample data (\mathbf{y}, \mathbf{X}) . In this scheme, the rows of the observed data matrix, namely the (y_i, \mathbf{x}_i') , are independent and identically distributed across i . So from Eq. (16.5), it follows that $E(y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$ and $V(y_i|\mathbf{x}_i) = \sigma^2$. But also $E(y_i) = \mu_y$ for all i , $V(y_i) = \sigma_y^2$ for all i , and the \mathbf{X} matrix is random. So this sampling scheme does not support the CR model, in which the expectations of the y_i differ and the \mathbf{X} matrix is not random.

16 Classical Regression: Interpretation and Application

16.1. Interpretation of the Classical Regression Model

It is instructive to compare our specification of the classical regression model to the more customary one. Our CR model is specified as

$$(16.1) \quad E(y) = \mathbf{X}\boldsymbol{\beta},$$

$$(16.2) \quad V(y) = \sigma^2 \mathbf{I},$$

$$(16.3) \quad \mathbf{X} \text{ nonstochastic,}$$

$$(16.4) \quad \text{rank}(\mathbf{X}) = k.$$

Judge et al. (1988, pp. 178–183) specify a "General Linear Statistical Model" as follows (notation has been slightly changed):

$$(16.1^*) \quad y = \mathbf{X}\boldsymbol{\beta} + \epsilon,$$

$$(16.2^*) \quad \mathbf{X} \text{ is a known nonstochastic matrix with linearly independent columns,}$$

$$(16.3^*) \quad E(\epsilon) = \mathbf{0},$$

$$(16.4^*) \quad E(\epsilon\epsilon') = \sigma^2 \mathbf{I}.$$

The two models are equivalent. Judge et al.'s ϵ is simply the *disturbance vector*, the deviation of the random vector y from its expectation $\mu = \mathbf{X}\boldsymbol{\beta}$. In that style, for a scalar random variable y with $E(y) = \mu$ and $V(y) = \sigma^2$, one might write $y = \mu + \epsilon$, $E(\epsilon) = 0$, $E(\epsilon^2) = \sigma^2$. There is no serious objection to doing so, except that it tends to give the disturbance a life of its own, rather than treating it as merely the deviation of a random variable from its expected value. Doing so may make one think

Instead of random sampling, we will rely on:

Stratified Sampling from the Multivariate Population. Here n values of the random vector \mathbf{x} are specified. These values, the \mathbf{x}_i ($i = 1, \dots, n$), define n subpopulations, or strata. In the i th subpopulation, or stratum, the pdf or pmf of the dependent variable is $g(y|\mathbf{x}_i)$, with $E(y|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta} = \mu_i$, say, and $V(y|\mathbf{x}_i) = \sigma^2$. A random drawing is made from each subpopulation. That is, y_1 is drawn from subpopulation 1, y_2 is drawn from subpopulation 2, and so on. The successive drawings are independent. In this scheme, the sampled y 's are not identically distributed; they are drawn from different subpopulations. The list of n selected \mathbf{x}_i vectors is maintained in repeated sampling, so the expectations of the successive y 's will depend only on i . We can then write $E(y_i)$ instead of $E(y|\mathbf{x}_i)$, and similarly we can write $V(y_i)$ instead of $V(y|\mathbf{x}_i)$. There is no need for all the \mathbf{x}_i 's to differ: the relevant requirement is that $\text{rank}(\mathbf{X}) = k$, so we need k linearly independent (in the matrix-algebra sense) \mathbf{x}_i 's. As discussed in Section 13.5, stratified sampling does not require that the researcher control the \mathbf{x} values in the sense of imposing them on the subjects.

Under stratified sampling, it does not make sense to use the sample to estimate the population means and variances of the x 's and y . The sample on \mathbf{x} is not randomly drawn from the population joint distribution of \mathbf{x} , and consequently the sample on y is not randomly drawn from the population marginal distribution of y . Still, as in the bivariate case (Section 13.5), while stratification on \mathbf{x} does induce a new marginal distribution for \mathbf{x} and y , it preserves the conditional probability distributions of y given \mathbf{x} . That suffices when we are concerned with the conditional expectation of y given \mathbf{x} .

This stratified sampling scheme, also known as the nonstochastic explanatory variable scheme, will support the CR model. We adopt it now in order to simplify the theory. In Chapter 25 we will see how the conclusions carry over to the more natural scheme of random sampling.

Setting aside the sampling aspects, it is useful to compare this discussion of the underlying assumptions of the CR model with that in other textbooks. Johnston (1984, p. 169) seems to say that for the CR model to be correct, the theorist must have "done a good job in specifying all the significant explanatory variables." Judge et al. (1988, p. 186) say that "it is assumed that the \mathbf{X} matrix contains the correct set of explanatory variables. In real-world situations we seldom, if ever, know the correct set of explanatory variables, and, consequently, certain relevant

variables may be excluded or certain extraneous variables may be included." Here "the correct set of explanatory variables" seems to mean the variables that "could have or actually determined the outcomes that we have observed" (ibid., p. 178).

Such requirements are very stringent, and have a causal flavor that is not part of the explicit specification of the CR model. An alternative position is less stringent and is free of causal language. Nothing in the CR model itself requires an exhaustive list of the explanatory variables, nor any assumption about the direction of causality. We have in mind a joint probability distribution, in which any conditional expectation function is conceivably of interest. For example, suppose that the random vector (y, x_2, x_3) has a trivariate probability distribution. On the one hand, we might be interested in $E(y|x_2, x_3)$, but on the other hand we might be interested in $E(y|x_3)$ or, for that matter, in $E(x_3|x_2, y)$. It is possible that all of those CEF's are linear, and that none of them is causal. It may be true that causal relations are the most interesting ones, but that is a matter of economics rather than of statistics. More on this in Chapter 31.

16.2. Estimation of Linear Functions of $\boldsymbol{\beta}$

In the CR model we deal with an $n \times 1$ random vector y . In general such a vector would have $E(y) = \boldsymbol{\mu}$ and $V(y) = \boldsymbol{\Sigma}$. One thing that makes the CR model special is the assumption that $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, that is, $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$. The n unknown μ_i 's may well be distinct, but all of them are expressible in terms of only k unknown β_j 's.

In the CR model, we estimate $\boldsymbol{\beta}$ by \mathbf{b} , and thus estimate $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ by $\hat{\boldsymbol{\mu}} = \mathbf{X}\mathbf{b} = \mathbf{N}\mathbf{y} = \hat{\mathbf{y}}$, rather than by y itself. Now $E(\hat{\mathbf{y}}) = \boldsymbol{\mu}$, and also $E(y) = \boldsymbol{\mu}$, so both the fitted-value vector and the observed vector are unbiased estimators of $\boldsymbol{\mu}$. Why is it preferable to use $\hat{\mathbf{y}}$? An answer runs as follows. Because $V(y) = \sigma^2\mathbf{I}$ and $V(\hat{\mathbf{y}}) = \sigma^2\mathbf{N}$, we have

$$V(y) - V(\hat{\mathbf{y}}) = \sigma^2(\mathbf{I} - \mathbf{N}) = \sigma^2\mathbf{M}.$$

The matrix $\mathbf{M} = \mathbf{M}'\mathbf{M}$ is nonnegative definite, so $V(y) \geq V(\hat{\mathbf{y}})$.

With respect to a single element of $\boldsymbol{\mu}$, say μ_i , the preferred estimator is $\hat{\mu}_i = \mathbf{x}_i'\mathbf{b} = \hat{y}_i = \mathbf{n}_i'\mathbf{y}$, where \mathbf{n}_i' denotes the i th row of \mathbf{N} . A simpler unbiased estimator is $\bar{y}_i = \mathbf{h}_i'\mathbf{y}$, where \mathbf{h}_i is the $n \times 1$ vector with a 1 in its i th slot and zeroes elsewhere. Observe that $\hat{\mu}_i$ is a linear function of

all n of the y 's, while y_i is a function of only one of them. Evidently in the CR model it is desirable to combine information from all the observations in order to estimate the expectation of a single one. Such a preference is clear in random sampling from a univariate population, where all the observations have the same expectation. In the CR model, the preference persists even though the expectations are not the same. The reason is that the expectations are linked together, being functions of the same k β_j 's.

Now, $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$ is a special case of a linear combination of the β 's. The general case is $\theta = \mathbf{h}'\boldsymbol{\beta}$ where \mathbf{h} is a nonrandom $k \times 1$ vector. Other special cases are of interest. For example, take $\mathbf{h} = (0, 0, 1, 0, \dots, 0)'$, then $\theta = \beta_3$; or take $\mathbf{h} = (0, 1, -1, 0, \dots, 0)'$, then $\theta = \beta_2 - \beta_3$. As indicated in Section 15.4, the preferred estimator of such a θ in the CR model is $t = \mathbf{h}'\mathbf{b}$. We elaborate on that point here.

By linear function rules, $E(t) = \mathbf{h}'E(\mathbf{b}) = \mathbf{h}'\boldsymbol{\beta} = \theta$, so that t is an unbiased estimator of θ . Further, $V(t) = \mathbf{h}'V(\mathbf{b})\mathbf{h} = \sigma^2\mathbf{h}'\mathbf{Q}^{-1}\mathbf{h}$. We can express t as a linear function of \mathbf{y} : $t = \mathbf{h}'\mathbf{b} = \mathbf{h}'\mathbf{A}\mathbf{y} = \mathbf{w}'\mathbf{y}$, where $\mathbf{w} = \mathbf{A}'\mathbf{h}$ is $n \times 1$ and nonstochastic. Consider all linear functions of \mathbf{y} that might be used to estimate θ : $t^* = \mathbf{w}^*\mathbf{y}$, where \mathbf{w}^* is a nonstochastic $n \times 1$ vector. We have

$$E(t^*) = \mathbf{w}^*\boldsymbol{\mu} = \mathbf{w}^*\mathbf{X}\boldsymbol{\beta}, \quad V(t^*) = \mathbf{w}^*\boldsymbol{\Sigma}\mathbf{w}^* = \sigma^2\mathbf{w}^{*'}\mathbf{w}^*,$$

so t^* is unbiased iff $\mathbf{w}^*\mathbf{X} = \mathbf{h}'$. In that event, we can write

$$\mathbf{h}'\mathbf{Q}^{-1}\mathbf{h} = \mathbf{w}^*\mathbf{X}\mathbf{Q}^{-1}\mathbf{X}'\mathbf{w}^* = \mathbf{w}^{*'}\mathbf{N}\mathbf{w}^*,$$

and thus write

$$V(t) = \sigma^2\mathbf{h}'\mathbf{Q}^{-1}\mathbf{h} = \sigma^2\mathbf{w}^{*'}\mathbf{N}\mathbf{w}^*.$$

Observe that

$$V(t^*) - V(t) = \sigma^2\mathbf{w}^{*'}(\mathbf{I} - \mathbf{N})\mathbf{w}^* = \sigma^2\mathbf{w}^{*'}\mathbf{M}\mathbf{w}^* \geq 0,$$

because \mathbf{M} is nonnegative definite. Thus the natural estimator of $\theta = \mathbf{h}'\boldsymbol{\beta}$, namely $t = \mathbf{h}'\mathbf{b}$, is in fact MVLUE in the CR model, where "linear" means linear in \mathbf{y} .

In practice, we will want to give some indication of the reliability of our estimate of θ . To estimate $V(t)$, replace σ^2 by $\hat{\sigma}^2$. The resulting standard error for $t = \mathbf{h}'\mathbf{b}$ is $\hat{\sigma}_t = \hat{\sigma}\sqrt{\mathbf{h}'\mathbf{Q}^{-1}\mathbf{h}}$.

To recapitulate: in the CR model, whether our interest is in estimating the full vector $\boldsymbol{\beta}$, in estimating one of its elements β_j , or in estimating a linear combination θ of its elements, the preferred procedure is to use LS linear regression.

16.3. Estimation of Conditional Expectation, and Prediction

In the CR model, to estimate the parameter $\mu_i = E(y_i) = \mathbf{x}_i'\boldsymbol{\beta}$, our conclusion was to use $\hat{y}_i = \mathbf{x}_i'\mathbf{b}$. The expectation and variance of this estimator are

$$E(\hat{y}_i) = \mu_i, \quad V(\hat{y}_i) = \sigma^2\mathbf{x}_i'\mathbf{Q}^{-1}\mathbf{x}_i = \sigma^2\eta_i,$$

where η_i is the i th diagonal element of \mathbf{N} . Now suppose that we are interested in estimating a point on the CEF, say $\mu_0 = \mathbf{x}_0'\boldsymbol{\beta}$, where \mathbf{x}_0 is some $k \times 1$ vector, not necessarily one of the points at which we have sampled. This parameter μ_0 is the expectation of y_0 , where y_0 is a random drawing from the subpopulation defined by $\mathbf{x} = \mathbf{x}_0$. Because μ_0 is a linear combination of the elements of $\boldsymbol{\beta}$, the preferred estimator for it is $\hat{\mu}_0 = \mathbf{x}_0'\mathbf{b}$, which has expectation and variance

$$E(\hat{\mu}_0) = \mu_0, \quad V(\hat{\mu}_0) = \sigma^2\mathbf{x}_0'\mathbf{Q}^{-1}\mathbf{x}_0.$$

The standard error for this estimator is $\hat{\sigma}\sqrt{\mathbf{x}_0'\mathbf{Q}^{-1}\mathbf{x}_0}$.

Prediction, or forecasting, is a distinct problem. There the objective is to predict the value of y_0 , a single random drawing from the subpopulation defined by $\mathbf{x} = \mathbf{x}_0$. If we knew $\boldsymbol{\beta}$, our prediction would be $\mu_0 = \mathbf{x}_0'\boldsymbol{\beta}$. The prediction error would be $\epsilon_0 = y_0 - \mu_0$, with expectation $E(\epsilon_0) = 0$ and variance $E(\epsilon_0^2) = V(y_0) = \sigma^2$. In practice, we do not know $\boldsymbol{\beta}$, but we have a sample from the CR model, from which we have calculated \mathbf{b} . The natural predictor will be $\hat{\mu}_0 = \mathbf{x}_0'\mathbf{b}$. When that predictor is used, the prediction error will be $u = y_0 - \hat{\mu}_0$, with

$$\begin{aligned} E(u) &= E(y_0) - E(\hat{\mu}_0) = 0, \\ V(u) &= V(y_0) + V(\hat{\mu}_0) - 2C(y_0, \hat{\mu}_0) = \sigma^2 + \sigma^2\mathbf{x}_0'\mathbf{Q}^{-1}\mathbf{x}_0 \\ &= \sigma^2(1 + \mathbf{x}_0'\mathbf{Q}^{-1}\mathbf{x}_0), \end{aligned}$$

taking the covariance to be zero on the understanding that the drawing on y_0 is independent of the sample observations \mathbf{y} . This predictor is unbiased, and the variance of the prediction error has two additive

components: the variance of the prediction error that would be made were μ_0 known and used, and the variance of the estimator of μ_0 . The "standard error of forecast," which is the square root of the estimate of $V(w)$, is given by $\hat{\sigma}\sqrt{(1 + \mathbf{x}_0'Q^{-1}\mathbf{x}_0)}$.

16.4. Measuring Goodness of Fit

In empirical research that relies on the CR model, the objective is to estimate the population parameter vector β , rather than to "fit the data" or to "explain the variation in the dependent variable." Nevertheless it is customary to report, along with the parameter estimates and their standard errors, a measure of goodness of fit.

To develop the measure, return to the algebra of least squares. Given the data \mathbf{y} , $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, we have run the LS linear regression of \mathbf{y} on \mathbf{X} , obtaining the coefficient vector \mathbf{b} , the fitted-value vector $\hat{\mathbf{y}}$, and the residual vector \mathbf{e} . Observe that $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, and that $\hat{\mathbf{y}}'\mathbf{e} = (\mathbf{Xb})'\mathbf{e} = \mathbf{b}'(\mathbf{X}'\mathbf{e}) = \mathbf{b}'\mathbf{0} = 0$, by the FOC's. So

$$\mathbf{y}'\mathbf{y} = (\hat{\mathbf{y}} + \mathbf{e})'(\hat{\mathbf{y}} + \mathbf{e}) = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e},$$

which algebraically is

$$(16.6) \quad \sum_i y_i^2 = \sum_i \hat{y}_i^2 + \sum_i e_i^2.$$

This is an *analysis* (that is, decomposition) of *sum of squares*: the sum of squares of observed values is equal to the sum of squares of the fitted values plus the sum of squares of the residuals.

Further, $\sum_i y_i = \sum_i \hat{y}_i + \sum_i e_i$, so the mean of the observed values equals the mean of the fitted values plus the mean of the residuals:

$$\bar{y} = \bar{\hat{y}} + \bar{e}.$$

Now if $\bar{e} = 0$, then $\bar{y} = \bar{\hat{y}}$, so $n\bar{y}^2 = n\bar{\hat{y}}^2$, which subtracted from the decomposition in Eq. (16.6) gives

$$(16.7) \quad \sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i e_i^2.$$

This is an *analysis of variation*, where variation is defined to be the sum of squared deviations about the sample mean. Provided that the mean residual is zero, the variation of the observed values is equal to the variation of the fitted values plus the variation of the residuals.

Divide Eq. (16.7) through by $\sum_i (y_i - \bar{y})^2$ to get

$$(16.8) \quad R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2}.$$

The measure R^2 , which will lie between zero and unity, is called the *coefficient of determination*, or squared multiple correlation coefficient. It measures, one says, the proportion of the variation of y that is accounted for (linearly) by variation in the x 's; note that the fitted value \hat{y}_i is an exact linear function of the x_i 's. In this sense, R^2 measures the goodness of fit of the regression.

Consider an extreme case:

$$R^2 = 1 \Leftrightarrow \sum_i e_i^2 = 0 \Leftrightarrow \mathbf{e}'\mathbf{e} = 0 \Leftrightarrow \mathbf{e} = \mathbf{0} \Leftrightarrow \mathbf{y} = \mathbf{Xb},$$

in which case the observed y 's fall on an exact linear function of the x 's. The fit is perfect; all of the variation in y is accounted for by the variation in the x 's. At the other extreme:

$$R^2 = 0 \Leftrightarrow \sum_i (\hat{y}_i - \bar{y})^2 = 0 \Leftrightarrow \hat{y}_i = \bar{y} \quad \text{for all } i,$$

in which case the best-fitting line is horizontal, and none of the variation in y is accounted for by variation in the x 's.

From our perspective, R^2 has a very modest role in regression analysis, being a measure of the goodness of fit of a sample LS linear regression in a body of data. Nothing in the CR model requires that R^2 be high. Hence a high R^2 is not evidence in favor of the model, and a low R^2 is not evidence against it. Nevertheless, in empirical research reports, one often reads statements to the effect that "I have a high R^2 , so my theory is good," or "My R^2 is higher than yours, so my theory is better than yours."

In fact the most important thing about R^2 is that it is not important in the CR model. The CR model is concerned with parameters in a population, not with goodness of fit in the sample. In Section 6.6 we did introduce the population coefficient of determination ρ^2 , as a measure of strength of a relation in the population. But that measure will not be invariant when we sample selectively, as in the CR model, because it depends upon the marginal distribution of the explanatory variables. If one insists on a measure of predictive success (or rather failure), then $\hat{\sigma}^2$ might suffice: after all, the parameter σ^2 is the expected squared

forecast error that would result if the population CEF were used as the predictor. Alternatively, the squared standard error of forecast (Section 16.3) at relevant values of \mathbf{x} may be informative.

Some further remarks on the coefficient of determination follow.

- One should not calculate R^2 when $\bar{e} \neq 0$, for then the equivalence of the two versions of R^2 in Eq. (16.8) breaks down, and neither of them is bounded between 0 and 1. What guarantees that $\bar{e} = 0$? The only guarantee can come from the FOC's $\mathbf{X}'\mathbf{e} = \mathbf{0}$. It is customary to allow for an intercept in the regression, that is, to have, as one of the columns of \mathbf{X} , the $n \times 1$ vector $\mathbf{s} = (1, 1, \dots, 1)'$. We refer to this \mathbf{s} as the *summer vector*, because multiplying \mathbf{s}' into any vector will sum up the elements in the latter. If \mathbf{s} is one of the columns in \mathbf{X} , then $\mathbf{s}'\mathbf{e} = 0$ is one of the FOC's, so $\bar{e} = 0$. The same conclusion follows if there is a linear combination of the columns of \mathbf{X} that equals the summer vector. Also if \mathbf{y} and $\mathbf{x}_2, \dots, \mathbf{x}_k$ all have zero column means in the sample, then $\bar{e} = 0$. But otherwise a zero mean residual is sheer coincidence.

- We can always find an \mathbf{X} that makes $R^2 = 1$: take any n linearly independent $n \times 1$ vectors to form the \mathbf{X} matrix. Because such a set of vectors forms a basis for n -space, any $n \times 1$ vector \mathbf{y} will be expressible as an exact linear combination of the columns of that \mathbf{X} . But of course "fitting the data" is not a proper objective of research using the CR model.

- The fact that R^2 tends to increase as additional explanatory variables are included leads some researchers to report an *adjusted* (or "corrected") *coefficient of determination*, which discounts the fit when k is large relative to n . This measure, referred to as \bar{R}^2 (read as " R bar squared"), is defined via

$$1 - \bar{R}^2 = (n - 1)(1 - R^2)/(n - k),$$

which inflates the unexplained proportion and hence deflates the explained proportion. There is no strong argument for using this particular adjustment: for example, $(1 - k/n)R^2$ would have a similar effect. It may well be preferable to report R^2 , n , and k , and let readers decide how to allow for n and k .

- The adjusted coefficient of determination may be written explicitly as

$$(16.9) \quad \bar{R}^2 = 1 - \left[\sum_i e_i^2 / (n - k) \right] / \left[\sum_i (y_i - \bar{y})^2 / (n - 1) \right].$$

It is sometimes said that in the CR model, the numerator $\sum_i e_i^2 / (n - k)$ is an unbiased estimator of the disturbance variance, and that the denominator $\sum_i (y_i - \bar{y})^2 / (n - 1)$ is an unbiased estimator of the variance of y . The first claim is correct, as we know. But the second claim is not correct: in the CR model the variance of the disturbance is the same thing as the common variance of the y_i , namely σ^2 .

Exercises

16.1 Continuing the numerical example of Exercises 14.1 and 15.6, assume that the CR model applies. Let $\theta = \beta_1 + \beta_2$. Report your estimate of θ , along with its standard error.

16.2 The CR model applies to $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ with $\sigma^2 = 1$. Here \mathbf{X} is an $n \times 2$ matrix with

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 4 & -2 \\ -2 & 3 \end{pmatrix}.$$

You are offered the choice of two jobs: estimate $\beta_1 + \beta_2$, or estimate $\beta_1 - \beta_2$. You will be paid the dollar amount $10 - (t - \theta)^2$, where t is your estimate and θ is the parameter combination that you have chosen to estimate. To maximize your expected pay, which job should you take? What pay will you expect to receive?

16.3 In a regression analysis of the relation between earnings and various personal characteristics, a researcher includes these explanatory variables along with six others:

$$x_7 = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male} \end{cases} \quad x_8 = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female} \end{cases}$$

but does not include a constant term.

- Does the sum of residuals from her LS regression equal zero?
- Why did she not also include a constant term?

16.4 Consider the customary situation, where the regression includes an intercept, and the first column of \mathbf{X} is $\mathbf{x}_1 = \mathbf{s}$, the summer vector. Let $\mathbf{M}_1 = \mathbf{I} - \mathbf{x}_1(\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'$.

personalizing it by completing names for the program and output files, and also entering your name. Run and print.

```

/* --1606 */ output file = --1606.OUT reset; format 8,4;
"Student name _____";
let x1 = 1 1 1 1; let x2 = 2 4 3 5 2; let y = 14 17 8 16 3;
X = x1~x2; Q = X'X; QI = invpd(Q); b = QI*X'y; sse = y'y - b'X'y;
"b' = " b' ; " sse = " sse; end;

```

(a) Show that $\mathbf{M}_1\mathbf{y}$ is the vector of residuals from a regression of y on the summer vector alone.

(b) Show that $\mathbf{y}'\mathbf{M}_1\mathbf{y} = \sum_i (y_i - \bar{y})^2$.

(c) Further suppose that the CR model applies to $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. Apply R5 (Section 15.1) to show that

$$E \left[\sum_i (y_i - \bar{y})^2 \right] = (n-1)\sigma^2 + \boldsymbol{\beta}_2' \mathbf{X}_2' \mathbf{X}_2 \boldsymbol{\beta}_2,$$

where $\boldsymbol{\beta}_2$ is the $(k-1) \times 1$ subvector that remains when the first element of $\boldsymbol{\beta}$ is deleted, \mathbf{X}_2 is the $n \times (k-1)$ submatrix that remains when the first column of \mathbf{X} is deleted, and $\mathbf{X}_2' = \mathbf{M}_1\mathbf{X}_2$.

(d) Evaluate the claim that in Eq. (16.9), the denominator of the adjusted coefficient of determination is an unbiased estimator of the variance of the dependent variable.

16.5 GAUSS is a mathematical and statistical programming language, produced by Aptech Systems, Inc., Kent, Washington. We will rely on it frequently in the remainder of this book, presuming that it is installed on a computer available to you. Appendix B provides some introductory information about GAUSS; other information will be provided as hints in subsequent exercises. Version 1.49B of GAUSS is used here; modification to other versions should be straightforward.

Here is a GAUSS program to re-do Exercise 14.1. Enter it, run it, and print out the program file and the output file.

```

/* ASG1605 */ output file = ASG1605.OUT reset; format 8,4;
let x1 = 1 1 1 1; let x2 = 2 4 3 5 2; let y = 14 17 8 16 3;
X = x1~x2; Q = X'X; dq = det(Q); QI = invpd(Q);
A = QI*X'; N = X*A; I = eye(5); M = I - N;
trm = sumc(diag(N)); trm = sumc(diag(M)); b = A*y; yh = N*y; e = M*y;
"Q = " Q; "det Q = " dq; ?;
"Q inverse = " QI; ?;
"N = " N; "M = " M; ?;
"tr(N) = " trm; "tr(M) = " trm; ?;
"b' = " b'; ?;
"yhat' = " yh'; ?; "e' = " e'; end;

```

16.6 The algorithm used in Exercise 16.5 is not an efficient way to run LS linear regressions. Here is a more sensible way, which may serve as a starting point for your own future regression programs. The program also calculates the sum of squared residuals. Enter the program,